SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner # :_____ Date: _____

Art Unit: _____ Phone Number 30_____ Serial Number: _____

Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.
****

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

_____

Earliest Priority Filing Date: _____

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

Point of Contact:
Jan Dalaval
Librarian-Physical Sciences
CM1 1E01 Tel: 308-4498

****

STAFF USE ONLY	Type of Search	Vendors and cost where applicable
Searcher: _____ Jan	NA Sequence (#)____ ✓	STN _____
Searcher Phone #: _____ 4458	AA Sequence (#)____ ✓	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: 1/8	Bibliographic _____	Dr.Link _____
Date Completed: 1/9	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: _____	Fulltext _____	Sequence Systems ____ ✓ ____
Clerical Prep Time: _____ 0	Patent Family _____	WWW/Internet _____
Online Time: _____ 0	Other _____	Other (specify)_____

PTO-1590 (8-01)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

protein - protein search, using sw model

Run on: January 8, 2002, 10:50:18 ; Search time 12.48 Seconds
            (without alignments)
            207.362 Million cell updates/sec

Query: 625
Perfect score: 625
Title: US-09-640-041-4
Sequence: 1 MPTDHEEPCGPSHKSFCLNG.........YDINLVETSSTSAHHSHEQH 115

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

database :

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Total number of hits satisfying chosen parameters: 212252

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

suit No.	Score	% Query Match	Length	DB	ID	Description
1	142	22.7	713	3	US-08-899-437-1	Sequence 2, Appli
2	142	22.7	713	1	US-09-126-121-2	Sequence 2, Appli
3	140	22.4	602	1	US-08-168-091A-2	Sequence 2, Appli
4	140	22.4	602	1	US-08-428-926-5	Sequence 5, Appli
5	140	22.4	602	1	US-08-428-927-5	Sequence 5, Appli
6	140	22.4	602	1	US-08-428-298-5	Sequence 5, Appli
7	140	22.4	602	1	US-08-339-517-5	Sequence 5, Appli
8	140	22.4	1070	1	US-08-899-954-3	Sequence 2, Appli
9	136	21.8	696	3	US-08-899-437-23	Sequence 23, Appli
10	136	21.8	696	3	US-09-126-121-23	Sequence 23, Appli
11	136	21.8	720	4	US-08-899-437-6	Sequence 6, Appli
12	136	21.8	720	4	US-09-126-121-6	Sequence 6, Appli
13	132.5	21.2	560	4	US-08-341-018-58	Sequence 58, Appli
14	132.5	21.2	560	4	US-08-470-335-194	Sequence 194, App
15	132.5	21.2	560	4	US-09-188-930-187	Sequence 187, App
16	130.5	20.9	152	4	US-07-847-743B-26	Sequence 227, App
17	130.5	20.9	604	4	US-08-470-335-227	Sequence 227, App
18	126.5	20.2	625	1	US-08-456-201-26	Sequence 26, Appl
19	126.5	20.2	625	2	US-08-456-241-26	Sequence 26, Appl
20	126.5	20.2	625	4	US-08-470-335-240	Sequence 240, App
21	126.5	20.2	625	1	US-08-456-241-26	Sequence 26, Appl
22	126.5	20.2	638	4	US-08-470-335-240	Sequence 26, Appl
23	126.5	20.2	625	2	PCT-US92-04295A-26	Sequence 26, Appl
24	126.5	20.2	625	1	US-07-847-743B-13	Sequence 13, Appl
25	126.5	20.2	669	1	US-08-456-201-8	Sequence 8, Appli
26	126.5	20.2	669	1	US-08-456-201-13	Sequence 13, Appl
27	126.5	20.2	669	2	US-08-330-161-11	Sequence 11, Appl
28	126.5	20.2	669	2	US-08-456-241-8	Sequence 8, Appli
29	126.5	20.2	669	2	US-08-456-241-13	Sequence 13, Appl
30	126.5	20.2	669	2	US-08-440-401-11	Sequence 11, Appl
31	126.5	20.2	669	2	US-08-419-878B-11	Sequence 11, Appl
32	126.5	20.2	669	5	PCT-US92-04295A-8	Sequence 8, Appli
33	126.5	20.2	669	5	PCT-US92-04295A-13	Sequence 13, Appl
34	126.5	20.2	821	4	US-08-470-335-228	Sequence 228, App
35	126.5	20.2	855	4	US-08-470-335-241	Sequence 241, App
36	126.5	20.2	868	4	US-08-470-335-229	Sequence 229, App
37	126.5	20.2	902	4	US-08-470-335-242	Sequence 242, App
38	124.5	19.9	121	1	US-07-847-743B-29	Sequence 14, Appl
39	124	19.8	420	1	US-08-549-757A-14	Sequence 29, Appl
40	124	19.8	420	1	US-07-847-743B-29	Sequence 29, Appl
41	124	19.8	420	2	US-08-456-201-29	Sequence 29, Appl
42	124	19.8	420	5	US-08-456-241-29	Sequence 29, Appl
43	124	19.8	601	4	PCT-US92-04295A-29	Sequence 29, Appl
44	124	19.8	635	4	US-08-470-335-233	Sequence 233, App
45	124	19.8	637	1	US-07-847-743B-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-08-899-437-2
Sequence 2, Application US/08899437
Patent No. 6121415

GENERAL INFORMATION:
APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
TITLE OF INVENTION: Ligands and Uses Therefor
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/899,437
FILING DATE: 24-Jul-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Conley, Deirdre L.
REGISTRATION NUMBER: 36,487
REFERENCE/DOCKET NUMBER: P1084R1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-2066
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 713 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
FEATURE:
NAME/KEY: Mouse NRG3 (mNRG3)/amino acid seq.
LOCATION: 1-713
IDENTIFICATION METHOD:
OTHER INFORMATION:
US-08-899-437-2

Query Match   22.7%; Score 142; DB 3; Length 713;
Best Local Similarity   26.6%; Pred. No. 6.1e-08;
Matches   38; Conservative 20; Mismatches 47; Indels 38; Gaps 6;

Qy 3 TDHEEPCGPSHKSFCLNGGLCYVIPTIPS--PFCRCVENYTGARCEEVFLP--------51

BASE COUNT   90 a   85 c   80 g   93 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Pred. No. 4.3e-103;
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 atgccaacagatcacgaagagccctgtggtcccagtcaacaagtcgttttgcctgaatggg   60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGCCAACAGATCACGAAGAGCCCTGTGGTCCCAGTCAACAAGTCGTTTTGCCTGAATGGG   60

Qy   61 gggcttttgttatgtgatacctacattcccagccccattttgtaggtcgttgaaaactat  120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGGCTTTTGTTATGTGATACCTACATTCCCAGCCCCATTTTGTAGGTCGTTGAAAACTAT  120

Qy  121 acaggagctcgttgtgaaggaggttttctctccagccccagccccatttttgtaggtcgttgaaaactaaccgtaac  180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 ACAGGAGCTCGTTGTGAAGGAGGTTTTCTCTCCAGCCCCAACCCAAACTAAAAGTAAC  180

Qy  181 ctgtttgaagcttttgtggcattggcggtcctagtaacacttattcattggagccttctac  240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CTGTTTGAAGCTTTTGTGGCATTGGCGGTCCTAGTAACACTTATTGTAGGCCTTCTAC  240

Qy  241 ttcctttgcaggaaagcccacttcagagagagcccagtccagtgtatgatcaacctg  300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 TTCCTTTGCAGGAAAGCCCACTTCAGAGAGAGCCCAGTTCAGTGTATGATATCAACCTG  300

Qy  301 gtagagacgagcagtaccagtgcccaccacagtcatgaacaacactga  348
        ||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 GTAGAGACGAGCAGTACCAGTGCCCACCACAGTCATGAACAACACTGA  348
```

RESULT 2
AX088170
LOCUS   AX088170   348 bp   DNA   PAT   17-MAR-2001
DEFINITION Sequence 1 from Patent WO0114415.
ACCESSION   AX088170
VERSION   AX088170.1 GI:13397083
KEYWORDS
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 348)
  AUTHORS   Kavanaugh,W.M., Cen,H. and Lee,P.
  TITLE   Egfh2 genes and gene products
  JOURNAL   Patent: WO 0114415-A 1 01-MAR-2001;
            CHIRON CORPORATION (US)
FEATURES   Location/Qualifiers
     source   1..348
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"

BASE COUNT   82 a   102 c   87 g   77 t
ORIGIN

Query Match
Best Local Similarity 82.8%; Pred. No. 1.2e-71;
Matches 288; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```
Qy    1 atgccaacagatcacgaagagccctgtggtcccagtcaacaagtcgttttgcctgaatggg   60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGCCAACAGATCACGAGCAGCGCCCTGTGGTCCCAGGTCAGGTCATTTTTGCCTCAATGGG   60

Qy   61 gggcttttgttatgtgatacctacattcccagccccattttgtaggtcgttgaaaactat  120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGGATTTGTTATGTGATACTACATTCCCAGCCCCATTTTGTAGGTCGTTGAAAACTAT  120

Qy  121 acaggagctcgttgtgaaggaggttttctctccagccccagccccattttgtaggtcgttgaaaactaaccgtaac  180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 ACCGGAGCACGCTCGGAAGAGAGGTTTTTCTCTCCCAAGCTCCAAGAAAGTAAT  180
```

RESULT 3
AF083067
LOCUS   AF083067   750 bp   mRNA   ROD   03-JUN-1999
DEFINITION Mus musculus neuregulin-4 short isoform (Nrg4) mRNA, complete cds.
ACCESSION   AF083067
VERSION   AF083067.2 GI:4539680
KEYWORDS
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 750)
  AUTHORS   Harari,D., Tzahar,E., Romano,J., Shelly,M., Pierce,J.H.,
            Andrews,G.C. and Yarden,Y.
  TITLE   Neuregulin-4: a novel growth factor that acts through the ErbB-4
            receptor tyrosine kinase
  JOURNAL   Oncogene 18 (17), 2681-2689 (1999)
  MEDLINE   99276098
   PUBMED   10348342
REFERENCE   2 (bases 1 to 750)
  AUTHORS   Harari,D. and Yarden,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (09-AUG-1998) Biological Regulation, The Weizmann
            Institute of Science, Rehovot 76100, Israel
COMMENT   On Mar 30, 1999 this sequence version replaced gi:4530129.
FEATURES   Location/Qualifiers
     source   1..750
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
     gene   1..750
                     /gene="Nrg4"
     CDS   64..411
                     /gene="Nrg4"
                     /note="ErbB-4 ligand"
                     /codon_start=1
                     /product="neuregulin-4 short isoform"
                     /protein_id="AAD21874.1"
                     /db_xref="GI:4530130"
                     /translation="MPTDHEQPCGPRHRSFCLNGGICYVIPTISPFCRCIENYTGAR
                     CEEVFLPSSSIPSESNLSAAFVVLAVLLTIIAACFLCRKGHLQRASSVQCEISLVE
                     TNNRTRHSHREH"

BASE COUNT   209 a   191 c   164 g   186 t
ORIGIN

Query Match
Best Local Similarity 82.8%; Pred. No. 1.4e-71;
Matches 288; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```
Qy    1 atgccaacagatcacgaagagccctgtggtcccagtcaacaagtcgttttgcctgaatggg   60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   64 ATGCCAACAGATCACGAGCAGCCCTGTGGTTCCCAGGCACAGGTCATTTTTGCCTCAATGGG  123

Qy   61 gggcttttgttatgtgatacctacattcccagccccattttgtaggtcgttgaaaactat  120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  124 GGGATTTGTTATGTGATACCTACTATCCCACAGCCCCATTTTGTAGGTCGTTGAAAATTAC  183

Qy  121 acaggagctcgttgtgaaggaggttttctctcccaggctccagcatccaaactaaaagtaac  180
        ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       acaggagctcgttgtgaaggaggttttctctcccaggctccagcatccaaactaaaagtaac  180
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

...protein - protein search, using sw model

...on: January 8, 2002, 10:49:43 ; Search time 23.58 Seconds
(without alignments)
361.257 Million cell updates/sec

...le: us-09-640-041-4

...fect score: 625

...quence: 1 MPTDHEEPCGPSHKSFCLNG...........YDINLVETSTSTSAHHSHEQH 115

...ring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

...rched: 522463 seqs, 74073290 residues
Listing first 45 summaries

...cal number of hits satisfying chosen parameters: 522463

...imum DB seq length: 0

...ximum DB seq length: 2000000000

...st-processing: Minimum Match 0%
Maximum Match 100%

...abase : A_Geneseq_1101:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result NO.	Score	% Query Match	Length	DB	ID	Description
1	625	100.0	115	22	AAY97697	Human EGFH2 protei
2	489	78.2	115	22	AAY97698	Mouse EGFH2 protei
3	142	22.7	713	20	AAW97617	Mouse neuregulin r
4	140	22.4	602	15	AAR51705	Neurotrophic facto
5	140	22.4	602	18	AAW32900	Chicken neurotroph
6	140	22.4	1070	22	AAB48099	Amino acid sequenc
7	136	21.8	696	20	AAW97619	Human neuregulin r
8	136	21.8	720	20	AAY05452	Human heregulin-11
9	136	21.8	720	20	AAW97618	Human neuregulin r
10	132.5	21.2	560	15	AAR46916	Human heregulin r
11	132.5	21.2	560	15	AAR55657	GGF2BPP4. Bos tau
12	132.5	21.2	560	16	AAR67245	Bovine glial cell
13	132.5	21.2	560	17	AAW09370	Human neuregulin G
14	132.5	21.2	560	20	AAR98763	Glial growth facto
15	132.5	21.2	560	17	AAR87456	BPP4 glial growth
16	132.5	21.2	560	20	AAY26576	Amino acid sequenc
17	130.5	20.9	152	21	AAY76009	Murine TGF-alpha h
18	130.5	20.9	152	22	AAB55948	Skin cell protein,
19	130	20.8	675	19	AAW74504	Amino acid sequenc
20	130	20.8	675	19	AAW74497	Amino acid sequenc
21	129	20.6	675	19	AAW74503	Human heregulin-be
22	129	20.6	675	13	AAR71182	Human heregulin HR
23	128	20.5	637	13	AAY05451	Human heregulin-11
24	127	20.3	675	20	AAR29573	Human heregulin-be
25	127	20.3	157	13	AAR71189	Human heregulin HR
26	126.5	20.2	462	16	AAR68562	Human NDF-alpha2b
27	126.5	20.2	462	16	AAR68561	Human proNDF-alpha
28	126.5	20.2	625	20	AAY06634	Heregulin-alpha.
29	126.5	20.2	669	13	AAR29570	Human heregulin-al
30	126.5	20.2	669	21	AAY71171	Human heregulin HR
31	126	20.2	675	21	AAY71198	Human Heregulin HR
32	124.5	19.9	121	16	AAR64230	HCT-15 contg. hTG-
33	124	19.8	276	16	AAR68566	Human proNDF-beta2
34	124	19.8	420	13	AAR29575	Human heregulin-be
35	124	19.8	420	20	AAY06638	Heregulin-beta-2-1
36	124	19.8	420	21	AAY71175	Herugelin-beta-2.
37	124	19.8	420	20	AAY06636	Herugelin-beta-2.
38	124	19.8	637	21	AAY71173	Human Heregulin HR
39	124	19.8	675	19	AAW74500	Amino acid sequenc
40	124	19.8	675	21	AAY71180	Human Heregulin HR
41	124	19.8	675	16	AAY71188	Human Heregulin HR
42	124	19.8	675	21	AAY71192	Human Heregulin HR
43	124	19.8	675	21	AAY71199	Human Heregulin HR
44	123.5	19.8	298	22	AAB57747	Amino acid sequenc
45	123.5	19.8	422	14	AAR43651	Rat NRSF. Rattus

ALIGNMENTS

RESULT 1
AAY97697

ID   AAY97697 standard; Protein; 115 AA.
XX
AC   AAY97697;
XX
DT   19-JUN-2001 (first entry)
XX
DE   Human EGFH2 protein sequence.
XX
KW   EGFH2; epidermal growth factor H2; EGFH2 protein modulated disorder;
KW   cancer; peripheral neuropathy; amyotrophic lateral sclerosis; therapy;
KW   Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW   ischaemic stroke; brain injury; acute spinal cord injury; infection;
KW   nervous system injury; multiple sclerosis; dementia; epilepsy;
KW   peripheral nerve injury; acoustic trauma; human.
OS   Homo sapiens.
XX
PN   WO200114415-A2.
XX
PD   01-MAR-2001.
XX
PF   15-AUG-2000; 2000WO-US223326.
XX
PR   20-AUG-1999; 99US-0149986.
XX
PA   (CHIR) CHIRON CORP.
XX
PI   Kavanaugh WM, Cen H, Lee P;
XX
DR   WPI; 2001-211304/21.
DR   N-PSDB; AAA91294.
XX

Sfil (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGGCCGAGGCGGCCGACGCGTG-dT(30)BN-3' (where B = A,
C, G or N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."

BASE COUNT   233 a   182 c   174 g   199 t
ORIGIN

Query Match   76.4%;   Score 266;   DB 11;   Length 788;
Best Local Similarity   96.6%;   Pred. No. 5e-70;
Matches   336;   Conservative   0;   Mismatches   5;   Indels   7;   Gaps   6;

```
Qy    1 atgccaacagatcacgaagaagagcccctgtggtcccagtcaaactaaaagtaac     60
        ||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db  161 ATGCCAACAGATCACGAAGAAGAGCCCTGTGGTCCCAGTCACAAGTCG-TTTGCCTGAATGGG 219

Qy   61 gggcttgttgttatgtgatacctactattcccagcccattttgtaggtgcgttgaaaactat 120
        |||| |||||| ||||||||||||||||| |||||||  |||||||||| ||||||||||||
Db  220 GGGC-TTGTTATGTGATACCTACTACTATTCCCAGCCCA--TTTGTAGGTGCCTTGAAAACTAT 277

Qy  121 acaggagctcgttgtgaggaaggtgttttctcccccgggctccagcccaaactaaaagtaac 180
        ||||||||||||||| ||||| |||||||||| || |||||||| |||||||||| |||||
Db  278 ACAGGAGCTCGTTGTGTGAAGAGGTTTTTCTCCCAGGCTCCAGCATCCAAACTAAAGTAAC   337

Qy  181 ctgtgttgaagcttttgtgtggcattggcggttcctagtaacacacttatcattggagccttctac 240
        |||| ||||||||  |||||||| ||||||||||||||||||   ||||| ||||||||||||
Db  338 CTGTTTTGAAGCTTT--GTGGCATGGCGGTTCCTAGTAACACTTATCA-TGGAGCCTTCTAC   394

Qy  241 ttcccttgcaggaaagccacacttcagagagcagttcagtccagtgtatgataatccaacctg 300
        ||| |||||||||||| ||||  ||||| | ||  ||||||||||||||||| |||| |||
Db  395 TTCCTTGCAGGAAAGGCCACTTCAGGAGACCCACAGTTCAGTCCAGTGTATGTCATGATTCATCAA 453

Qy  301 gtaggagacgagcagtacccagtgtgcccaccacagtgcatgaacaacactga      348
        ||| |||||||||||||||  |||| ||||||||||||| |||||||||||||
Db  454 GTAGAGAGACGAGCAGTACCAGTGCCCACCACAGTCATGAACAACACTGA          501
```

RESULT 2

FEATURES   Location/Qualifiers
source   1..568
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /clone="IMAGE:679804"
    /clone_lib="Soares mouse NML"
    /tissue_type="liver"
    /lab_host="DH10B"
    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
    polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
    was primed with a Not I - oligo(dT) primer [5'
    TGTTACCAATCTGAAGTGGGAGGAGGCGGCCGCGAAATCTTTTTTTTTTTTTTTTT
    TTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors
    (Pharmacia), digested with Not I and cloned into the Not I
    and Eco RI sites of the modified pT7T3 vector. Library
    was constructed and normalized by Bento Soares and M.Fatima
    Bonaldo."

High quality sequence stop: 459.

BASE COUNT   147 a   151 c   129 g   141 t
ORIGIN

Query Match   72.4%;   Score 252;   DB 10;   Length 568;
Best Local Similarity   82.8%;   Pred. No. 8.2e-66;
Matches   288;   Conservative   0;   Mismatches   60;   Indels   0;   Gaps   0;

```
Qy    1 atgccaacagatcacgaagaagagcccctgtggtcccagtcaaactaaaagtaac     60
        ||||||||||||||||| ||||  |||||||||||||||  ||||| |||||||||
Db   80 ATGCCAACAGATCACGAGCACCCCCTGTGGTCCCAGCCCATTTGTGGTGGTTCAGTTTGGGG 139

Qy   61 gggcttgttgttatgtgatacctactattcccagcccattttgtaggtgcgttgaaaactat 120
        ||||||| || ||||||||||||||||||||||||||||||||||||||  |||||||||||
Db  140 GGGATTTGTTTATGTGATACCTACTACTATCCCCACGCCCATTCGTAGGTTGCGTTGAAAATTAC 199

Qy  121 acaggagctcgttgtgaggaaggtgttttctcccccgggctccagcccaaactaaaagtaac 180
        ||||||||| |||||| ||||| ||||||||||||| |||||||| |||||||||| |||||
Db  200 ACCGGAGCACCGTGCGGAAGAGGTTTTTCTCCCCAGGGTCCCACCATTCGTGAGGTGCAGTTCCA 259

Qy  181 ctgtgttgaagcttttgtgtggcattggcggttcctagtaacacacttatcattggagccttctac 240
        |||||||||| ||||  ||||  |||||||| ||||||||| |||||||| ||||  |||||||
Db  260 CTGTCGGCAGCTTTCGTGGGTGTCGGCGGTCCTCCCCACTCTTTACCATCGCGCGCTCTGC 319

Qy  241 ttcccttgcaggaaagccacacttcagagagcagttcagtccagtgtatgataatccaacctg 300
        ||||| |||||||| |||||||  |||||| |||||||||||||||||||||| ||| |||
Db  320 TTCCCTGCAGGAAGGCCACCTTCAGGAGGGCCAGTTCAGTGTCAGTGGAGATCAGCCTG 379

Qy  301 gtaggagacgagcagtacccagtgtgcccaccacagtgcatgaacaacactga      348
        ||| ||||||   |  ||||| | |||||||||||| |||||||||| |||||
Db  380 GTAGAGAGACAAACAATACCAGAACCCGTCACAGCCACAGAGAACACTGA          427
```

RESULT 3
BF061527/c

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

protein - protein search, using sw model

on: January 8, 2002, 10:58:38 ; Search time 22.24 Seconds
(without alignments)
756.354 Million cell updates/sec

protein on:

tle: us-09-640-041-4
rfect score: 625
quence: 1 MPFDHEEPCGPSHKSFCLNG.........YDINLVETSSTSAHHSHEQH 115

oring table: BLOSUM62
oring table: Gapop 10.0 , Gapext 0.5

arched: 473505 seqs, 146272329 residues

al number of hits satisfying chosen parameters: 473505

st-processing: Minimum Match 0%
ximum DB seq length: 2000000000
nimum DB seq length: 0
Maximum Match 100%
Listing first 45 summaries

tabase : SPTREMBL_17:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

sult NO.	Score	% Query Match	Length	DB	ID	Description
1	130.5	20.9	461	11	O35947	O35947 mesocricetu
2	128.5	20.6	1214	6	Q9DD00	Q9bdq0 sus scrofa
3	123.5	19.8	298	11	Q9ESA9	Q9esa9 rattus norv
4	123.5	19.8	695	11	Q9ESB0	Q9esb0 rattus norv
5	122.5	19.6	162	11	Q61521	Q61521 mus musculu
6	118.5	19.0	162	11	Q9Z0L5	Q9z0l5 rattus norv
7	117.5	18.8	127	12	Q99F3	Q99f3 ectromelia
8	116	18.6	700	11	Q9ESB1	Q9esb1 rattus norv
9	116	18.6	782	11	Q9ESA5	Q9esa5 rattus norv
10	115	18.4	140	12	O8756	08756 variola vir
11	113.5	18.2	127	12	O8756	08756 variola vir
12	113.5	18.2	140	12	O41506	041506 bean 58058
13	113.5	18.2	140	12	O57166	057166 vaccinia vi
14	112	17.9	140	12	Q9FH4	Q9fh4 vaccinia vi
15	111	17.8	140	12	Q89066	Q89066 variola vir
16	109.5	17.5	138	12	Q89066	Q89066 variola vir
17	109.5	17.5	127	12	O86607	086607 vaccinia vi
18	108.5	17.4	125	12	P87605	P87605 cowpox viru
19	107.5	17.2	111	12	Q07112	Q07112 bos taurus
					O41504	041504 cowpox viru
					Q9ESA8	Q9esa8 rattus norv
20	107.5	17.2	136	11	Q9ESA7	Q9esa7 rattus norv
21	107.5	17.2	256	11	Q9ESA6	Q9esa6 rattus norv
22	107.5	17.2	317	11	Q9ESA3	Q9esa3 rattus norv
23	107.5	17.2	323	11	Q9ESA2	Q9esa2 rattus norv
24	107.5	17.2	342	11	Q9ESA1	Q9esa1 rattus norv
25	107.0	17.0	642	13	P79941	P79941 xenopus lae
26	106.5	16.9	295	5	Q9VLL2	Q9vll2 drosophila
27	105.5	16.9	4599	4	Q9NZR2	Q9nzr2 homo sapien
28	105	16.8	212	13	Q9WC5	Q9wc5 gallus gall
29	105	16.8	4601	5	Q9V383	Q9v383 drosophila
30	103	16.5	247	6	Q9BDH4	Q9bdh4 sus scrofa
31	101.5	16.2	4599	11	Q9J118	Q9j118 mus musculu
32	100	16.0	707	5	Q9VRQ3	Q9vrq3 drosophila
33	99.5	15.9	551	13	O13003	O13003 gallus gall
34	98.5	15.8	159	4	Q9BS56	Q9bs56 homo sapien
35	98.5	15.8	161	4	Q9UKM1	Q9ukm1 homo sapien
36	98.5	15.8	162	4	Q9UKM2	Q9ukm2 homo sapien
37	98.5	15.8	163	4	Q9UKM3	Q9ukm3 homo sapien
38	98.5	15.4	188	11	Q9R0C7	Q9r0c7 cricetulus
39	96	15.4	1643	6	O77611	O77611 bos taurus
40	95	15.2	105	11	Q99MN7	Q99mn7 mesocricetu
41	95	15.2	746	6	Q9H1R1	Q9h1r1 homo sapien
42	95	15.0	89	6	O28867	028867 equus cabal
43	94	15.0	656	6	O77612	O77612 bos taurus
44	94	15.0	1372	5	P91526	P91526 caenorhabdi
45	94	15.0	2394	6	O77610	O77610 bos taurus

RESULT 1
O35947

PRELIMINARY;   PRT;   461 AA.

ALIGNMENTS

ID   O35947   PRELIMINARY;   PRT;   461 AA.
AC   O35947;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PRO-NEUREGULIN-1, ISOFORM ALPHA 2B PRECURSOR.
GN   NRG1 OR NDF.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM ALPHA2B), AND SEQUENCE OF 64-81.
RC   TISSUE=EMBRYO;
RX   MEDLINE=98196996; PubMed=9537646;
RA   Velasco J.A., Feijoo E., Avila M.A., Notario V.;
RT   "Secretion of neu differentiation factor-like polypeptides by cph-
RT   transformed fibroblasts: cloning and characterization of Syrian
RT   hamster neuregulin cDNAs.";
RL   Mol. Carcinog. 21:156-163(1998).
CC   -!- FUNCTION: DIRECT LIGAND FOR ERBB3 AND ERBB4 TYROSINE KINASE
CC   RECEPTORS. CONCOMITANTLY RECRUITS ERBB1 AND ERBB2 CORECEPTORS,
CC   RESULTING IN LIGAND-STIMULATED TYROSINE PHOSPHORYLATION AND
CC   ACTIVATION OF THE ERB RECEPTORS. MAY PLAY AN IMPORTANT ROLE IN
CC   PROVIDING GROWTH ADVANTAGE IN NEOPLASTIC CELLS.
CC   -!- SUBUNIT: THE CYTOPLASMIC DOMAIN INTERACTS WITH THE LIM DOMAIN
CC   REGION OF LIMK1 (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: EXISTS AS TYPE I MEMBRANE PROTEIN AND AS A
CC   PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE MEMBRANE-
CC   BOUND FORM DOES NOT SEEM TO BE ACTIVE (BY SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: DIFFERENT ISOFORMS ARE PRODUCED BY
CC   ALTERNATIVE SPLICING. THE SEQUENCE SHOWN IS THAT OF ISOFORM
CC   ALPHA2B/CLONE PM3.
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT HIGHER LEVEL AFTER NEOPLASMIC
CC   TRANSFORMATION OF CELLS.
CC   -!- DOMAIN: THE CYTOPLASMIC DOMAIN MAY BE INVOLVED IN THE REGULATION
CC   OF TRAFFICKING AND PROTEOLYTIC PROCESSING. REGULATION OF THE
CC   PROTEOLYTIC PROCESSING INVOLVES INITIAL INTRACELLULAR DOMAIN
CC   DIMERIZATION (BY SIMILARITY).

```
T   Kavanaugh WM, Cen H, Lee P;
X
X   WPI; 2001-211304/21.
R   P-PSDB; AAY97697.
X
X   Novel epidermal growth factor H2 (EGFH2) polypeptides and
T   polynucleotides encoding them useful for treating EGFH2
T   protein-modulated disorders such as cancers, Alzheimer's disease,
T   ischaemic stroke and epilepsy .
XX
'S   Claim 2: Fig 2: 48pp; English.
X
XC   This sequence encodes the human epidermal growth factor H2 (EGFH2)
XC   protein of the invention. The EGFH2 DNA and protein sequences are useful
XC   for modulating the amount of EGFH2 protein in a subject, or treating an
XC   EGFH2 protein modulated disorder. The DNA or its antisense construct,
XC   ribozyme or a retroviral vector comprising a promoter are also useful
XC   for treating breast, prostate, pancreatic, oral or ovarian cancers,
XC   peripheral neuropathy, amyotrophic lateral sclerosis, Alzheimer's
XC   disease, Parkinson's disease, Huntington's disease, ischaemic stroke,
XC   brain injury, acute spinal cord injury, nervous system injury, multiple
XC   sclerosis, infection, dementia, epilepsy, peripheral nerve injury,
XC   acoustic trauma and tissue wound. The DNA and protein are also useful for
XC   expressing them encapsulated in a patient by implanting cells
XC   providing trophic support for cells in a patient by implanting cells
XC   expressing them encapsulated in a semipermeable membrane. EGFH2
XC   polypeptides can be used to screen peptide libraries. Antibodies which
XC   bind to EGFH2 and/or variant polypeptides can be used in diagnosing and
XC   determining the prognosis of cancer, tumour progression,
XC   hyperproliferative cell growth or accompanying biological and physical
XC   manifestations.
XX
SQ   Sequence 348 BP; 90 A; 85 C; 80 G; 93 T; 0 other;

Query Match            100.0%;     Score 348;     DB 22;     Length 348;
Best Local Similarity   100.0%;    Pred. No. 2.6e-106;
Matches 348; Conservative   0; Mismatches   0;     Indels   0;     Gaps   0;

Qy      1 atgccaacagatcacgaagagccctgtggtcccaagtcacaagtcgttttgcctgaatgg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 atgccaacagatcacgaagagccctgtggtcccaagtcacaagtcgttttgcctgaatgg 60

Qy     61 gggcttgttatgtgatgataccctactactattcccagcccatttgtaggtgcgttgaaaactat 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 gggcttgttatgtgatgataccctactactattcccagcccatttgtaggtgcgttgaaaactat 120

Qy    121 acaggagctcgttgtgtgaagagggtttttctctcccaggctccagcatccaaactaaaagtaac 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 acaggagctcgttgtgtgaagagggtttttctctcccaggctccagcatccaaactaaaagtaac 180

Qy    181 ctgtttgaagcttttgttggcattggcggtcctagtaaccattatcattggagccttctac 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ctgtttgaagcttttgttggcattggcggtcctagtaaccattatcattggagccttctac 240

Qy    241 ttcctttgcaggaaaggccacttcaggagagacgtcagttagtgatatcaaactg 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ttcctttgcaggaaaggccacttcaggagagacgtcagttagtgatatcaaactg 300

Qy    301 gtagaagacgacgagtaccaagtgcccaccacacagtcaatgaacaacactga 348
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 gtagaagacgacgagtaccaagtgcccaccacacagtcaatgaacaacactga 348
```

```
KW   EGFH2; epidermal growth factor H2; EGFH2 protein modulated disorder;
KW   cancer; peripheral neuropathy; amyotrophic lateral sclerosis; therapy;
KW   Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW   ischaemic stroke; brain injury; acute spinal cord injury; infection;
KW   nervous system injury; multiple sclerosis; dementia; epilepsy;
KW   peripheral nerve injury; acoustic trauma; mouse; ds.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT
FT   CDS             1..348
FT                   /tag= a
FT                   /product= "EGFH2"
XX
PN   WO200114415-A2.
XX
PD   01-MAR-2001.
XX
PF   15-AUG-2000; 2000WO-US22326.
XX
PR   20-AUG-1999;  99US-0149986.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Kavanaugh WM, Cen H, Lee P;
XX
DR   WPI; 2001-211304/21.
DR   P-PSDB; AAY97698.
XX
PT   Novel epidermal growth factor H2 (EGFH2) polypeptides and
PT   polynucleotides encoding them useful for treating EGFH2
PT   protein-modulated disorders such as cancers, Alzheimer's disease,
PT   ischaemic stroke and epilepsy -
XX
PS   Claim 2: Fig 3: 48pp; English.
XX
CC   This sequence encodes the mouse epidermal growth factor H2 (EGFH2)
CC   protein of the invention. The EGFH2 DNA and protein sequences are useful
CC   for modulating the amount of EGFH2 protein in a subject, or treating an
CC   EGFH2 protein modulated disorder. The DNA or its antisense construct,
CC   ribozyme or a retroviral vector comprising a promoter are also useful
CC   for treating breast, prostate, pancreatic, oral or ovarian cancers,
CC   peripheral neuropathy, amyotrophic lateral sclerosis, Alzheimer's
CC   disease, Parkinson's disease, Huntington's disease, ischaemic stroke,
CC   brain injury, acute spinal cord injury, nervous system injury, multiple
CC   sclerosis, infection, dementia, epilepsy, peripheral nerve injury,
CC   acoustic trauma and tissue wound. The DNA and protein are also useful for
CC   providing trophic support for cells in a patient by implanting cells
CC   expressing them encapsulated in a semipermeable membrane. EGFH2
CC   polypeptides can be used to screen peptide libraries. Antibodies which
CC   bind to EGFH2 and/or variant polypeptides can be used in diagnosing and
CC   determining the prognosis of cancer, tumour progression,
CC   hyperproliferative cell growth or accompanying biological and physical
CC   manifestations.
XX
SQ   Sequence 348 BP; 82 A; 102 C; 87 G; 77 T; 0 other;

Query Match             72.4%;     Score 252;     DB 22;     Length 348;
Best Local Similarity   82.8%;     Pred. No. 2.8e-74;
Matches 288; Conservative   0; Mismatches 60;     Indels   0;     Gaps   0;

Qy      1 atgccaacagatcacgaagagccctgtggtcccaagtcacaagtcgttttgcctgaatgg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db      1 atgccaacagatcacgaagcagccctgtggtcccaagtcacaagtcattttgcctcaatggg 60

Qy     61 gggcttgttatgtgatgataccctactactattcccagcccatttgtaggtgcgttgaaaactat 120
          ||| ||||||||||||  |||| |||||||||||||||||||||||| |||||||||| ||
Db     61 gggattgttatgtgatcccctactactattcccagcccattctgtaggttgccattgaaaatac 120

Qy    121 acaggagctcgttgtgtgaagagggtttttctctcccaggctccagcatccaaactaaaagtaac 180
          || ||||| || || ||||||||||||||||||||||||||||||||| ||||| ||||
Db    121 accggagcacgctgcgcaagcagaggttttctctcccaggctccagcgatccaaagcgaaagtaat 180
```

protein - protein search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

n on: January 8, 2002, 10:58:58 ; Search time 11.71 Seconds
(without alignments)
360.073 Million cell updates/sec

st-processing:
rfect score: 625
quence: 1 MPTDHEEPCGPSHKSFCLNG.............YDINLVFTSSTSAHHSHEQH 115
tle: US-09-640-041-4
oring table: BLOSUM62 Gapop 10.0 , Gapext 0.5
arched: 100059 seqs, 36664827 residues
tal number of hits satisfying chosen parameters: 100059
nimum DB seq length: 0
ximum DB seq length: 2000000000
st-processing: Minimum Match 0%
              Maximum Match 100%
              Listing first 45 summaries
tabase : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	% Query Match	Length	Match DB	ID	Description
1	489	78.2	115	1	Q9wtx4	mus musculu NRG4_MOUSE
2	142	22.7	713	1	O35181	mus musculu NRG3_MOUSE
3	140	22.4	602	1	Q05199	gallus gall NRG1_CHICK
4	136	21.8	720	1	P56975	homo sapien NRG3_HUMAN
5	126.5	20.2	639	1	Q02297	h pro-neure NRG1_HUMAN
6	124.5	19.9	169	1	O14944	homo sapien EREG_HUMAN
7	119.5	19.1	1217	1	P01132	mus musculu EGF_MOUSE
8	119	19.0	756	1	P56974	mus musculu NRG2_MOUSE
9	116.5	18.6	1133	1	P07522	rattus norv EGF_RAT
10	115	18.4	140	1	P33804	variola vir GRFA_VARV
11	115	18.1	53	1	O09968	sus scrofa EGF_PIG
12	113	18.1	4655	1	P98164	homo sapien LRP2_HUMAN
13	113	17.9	850	1	O14511	homo sapien NRG2_HUMAN
14	112	17.9	868	1	O35569	rattus norv NRG2_RAT
15	112	17.5	296	1	Q14491	homo sapien SMDF_HUMAN
16	109.5	17.5	1207	1	P01133	homo sapien EGF_HUMAN
17	109	17.4	677	1	O93383	xenopus lae NRG1_XENLA
18	108.5	17.4	140	1	P43322	r pro-neure NRG1_RAT
19	108.5	17.2	662	1	P42287	drosophila GRK_DROME
20	107.5	16.9	294	1	P20494	vaccinia v1 GRFA_VACCV
21	107.5	16.9	142	1	P08072	vaccinia vi GRFA_VACCC
22	105.5	16.6	85	1	O01580	myxoma viru GRFA_MYXVL
23	105	16.4	208	1	O01580	sus scrofa HBGF_PIG
24	103.5	16.2	4543	1	P98157	gallus gall LRP1_CHICK
25	102.5	16.2	208	1	P08441	sus scrofa HBGF_PIG
26	102.5	16.2	80	1	Q94918	shope fibro GRFA_SFVKA
27	101.5	16.1	622	1	P98135	drosophila VEIN_DROME
28	101	16.0	133	1	Q06922	ovis aries TGFA_SHEEP
29	101	15.9	160	1	O09118	sus scrofa TGFA_PIG
30	100.5	15.9	208	1	Q99075	cercopithec HBGF_CERAE
31	100	15.9	208	1	P01135	homo sapien HBGF_HUMAN
32	99.5	15.8	160	1	O06186	homo sapien TGFA_HUMAN
33	99.5	15.6	208	1	P55244	mus musculu HBGF_MOUSE
34	96.5	15.4	208	1	Q06175	rattus norv HBGF_RAT
35	96	15.4	178	1	P35070	homo sapien BTC_HUMAN
36	95.5	15.3	178	1	Q9ttc5	bos taurus BTC_BOVIN
37	93	14.9	402	1	P45442	caenorhabdi LAG2_CAEEL
38	93	14.9	2139	1	P10040	drosophila CRB_DROME
39	92	14.7	746	1	Q16819	homo sapien MEPA_HUMAN
40	91.5	14.6	159	1	P48030	mus musculu TGFA_MOUSE
41	91.5	14.6	159	1	P01134	rattus norv TGFA_RAT
42	91	14.6	177	1	P05928	mus musculu BTC_MOUSE
43	90.5	14.5	248	1	P31955	mus musculu SDGF_MOUSE
44	90.5	14.5	252	1	P15514	homo sapien AMPR_HUMAN
45	90	14.4	862	1	Q28858	macaca neme PGCV_MACNE

RESULT 1
NRG4_MOUSE

ALIGNMENTS

ID   NRG4_MOUSE   STANDARD;   PRT;   115 AA.
AC   Q9WTX4;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PRO-NEUREGULIN-4, SHORT ISOFORM (PRO-NRG4) [CONTAINS: NEUREGULIN-4
DE   (NRG-4)].
GN   NRG4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99276098; PubMed=10348342;
RA   Harari D., Tzahar E., Romano J., Shelly M., Pierce J.H., Andrews G.C.,
RA   Yarden Y.;
RT   "Neuregulin-4: a novel growth factor that acts through the ErbB-4
RT   receptor tyrosine kinase.";
RL   Oncogene 18:2681-2689(1999).
CC   -!- FUNCTION: LOW AFFINITY LIGAND FOR THE ERBB4 TYROSINE KINASE
CC   RECEPTOR. CONCOMITANTLY RECRUITS ERBB1 AND ERBB2 CORECEPTORS,
CC   RESULTING IN LIGAND-STIMULATED TYROSINE PHOSPHORYLATION AND
CC   ACTIVATION OF THE ERBB RECEPTORS. DOES NOT BIND TO THE ERBB1,
CC   ERBB2 AND ERBB3 RECEPTORS.
CC   -!- SUBCELLULAR LOCATION: EXISTS AS AN TYPE I MEMBRANE PROTEIN AND AS
CC   A PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE
CC   MEMBRANE-BOUND FORM DOES NOT SEEM TO BE ACTIVE (BY SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS MAY BE PRODUCED BY
CC   ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PANCREAS; WEAKLY EXPRESSED
CC   IN MUSCLE.
CC   -!- DOMAIN: THE CYTOPLASMIC DOMAIN MAY BE INVOLVED IN THE REGULATION
CC   OF TRAFFICKING AND PROTEOLYTIC PROCESSING. REGULATION OF THE
CC   PROTEOLYTIC PROCESSING INVOLVES INITIAL INTRACELLULAR DOMAIN
CC   DIMERIZATION (BY SIMILARITY).
CC   -!- DOMAIN: ERBB RECEPTOR BINDING IS ELICITED ENTIRELY BY THE EGF-LIKE
CC   DOMAIN (BY SIMILARITY).
CC   -!- PTM: PROTEOLYTIC CLEAVAGE CLOSE TO THE PLASMA MEMBRANE ON THE
CC   EXTERNAL FACE LEADS TO THE RELEASE OF THE SOLUBLE GROWTH FACTOR
CC   FORM (BY SIMILARITY).
CC   -!- PTM: EXTENSIVE GLYCOSYLATION PRECEDES THE PROTEOLYTIC CLEAVAGE (BY
CC   SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NEUREGULIN FAMILY.
CC   --

ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/899,437
FILING DATE: 24-Jul-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Conley, Deirdre L.
REGISTRATION NUMBER: 36,487
REFERENCE/DOCKET NUMBER: P1084R1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-2066
TELEFAX: 650/952-9881
REFERENCE/DOCKET NUMBER: P1084R1
SEQUENCE CHARACTERISTICS:
LENGTH: 2538 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
FEATURE:
NAME/KEY: mouse NRG3 nucleic acid
LOCATION: 1-2538
IDENTIFICATION METHOD:
OTHER INFORMATION:
US-08-899-437-1

Query Match
Best Local Similarity 55.0%;   Score 29.8; DB 3; Length 2538;
Matches 83; Conservative 0; Mismatches 62; Indels 6; Gaps 1;   Pred. No. 1.3;

Qy   4 ccaacagatcacgaagagccctgtggtcccagtcacaagtcgtttgcctgaatggggg 63
          | | ||||| | |||| ||||||| |||||| | ||||||||| || |||
Db 1141 CGATCTGAGCACTTCAAACCCTGTCGAGACAAGGACCTGGCGTATTGTCTCAATGATGGT 1200

Qy   64 ctttgttatgtgatacctactattcccagcccatt-----tgtaggtgcgttgaaaac 117
          | ||| ||| || ||| |||| || |||| || ||||| ||| ||| |||
Db 1201 GAATGCTTGTGATTGAGACACCCTGACAGGATCCCATAAGCACTGTCGGGTGCAAGGAAGGC 1260

Qy   118 tatacaggagcaccgttgtggaagaggttttc 148
          ||||| |||| |||||||||| ||| || |
Db 1261 TACCAAGGAGTCCGTTGTGATCAATTTCTGC 1291

RESULT 3
US-09-126-121-1
Sequence 1, Application US/09126121
Patent No. 6252051
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
TITLE OF INVENTION: Ligands and Uses Therefor
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WinPatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/126,121
FILING DATE: 30-Jul-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Conley, Deirdre L.
REGISTRATION NUMBER: 36,487
REFERENCE/DOCKET NUMBER: P1084R1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-2066
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2538 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
FEATURE:
NAME/KEY: mouse NRG3 nucleic acid
LOCATION: 1-2538
IDENTIFICATION METHOD:
OTHER INFORMATION:
US-09-126-121-1

Query Match
Best Local Similarity 55.0%;   Score 29.8; DB 4; Length 2538;
Matches 83; Conservative 0; Mismatches 62; Indels 6; Gaps 1;   Pred. No. 1.3;

Qy   4 ccaacagatcacgaagagccctgtggtcccagtcacaagtcgtttgcctgaatggggg 63
          | | ||||| | |||| ||||||| |||||| | ||||||||| || |||
Db 1141 CGATCTGAGCACTTCAAACCCTGTCGAGACAAGGACCTGGCGTATTGTCTCAATGATGGT 1200

Qy   64 ctttgttatgtgatacctactattcccagcccatt-----tgtaggtgcgttgaaaac 117
          | ||| ||| || ||| |||| || |||| || ||||| ||| ||| |||
Db 1201 GAATGCTTGTGATTGAGACACCCTGACAGGATCCCATAAGCACTGTCGGGTGCAAGGAAGGC 1260

Qy   118 tatacaggagcaccgttgtggaagaggttttc 148
          ||||| |||| |||||||||| ||| || |
Db 1261 TACCAAGGAGTCCGTTGTGATCAATTTCTGC 1291

RESULT 4
US-08-482-918-47
Sequence 47, Application US/08482918
Patent No. 6207417
GENERAL INFORMATION:
APPLICANT: Zsebo, Krisztina M.
APPLICANT: Bosselman, Robert A.
APPLICANT: Suggs, Sidney V.
APPLICANT: Martin, Francis H.
TITLE OF INVENTION: Stem Cell Factor
NUMBER OF SEQUENCES: 104
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,918
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107

GenCore version 4.5

protein - protein search, using sw model

n on: January 8, 2002, 10:56:28 ; Search time 14.34 Seconds
(without alignments)
610.884 Million cell updates/sec

tle: US-09-640-041-4
rfect score: 625
quence: 1 MPTDHEEPCGPSHKSFCLNG..........YDINLVETSSTSAHHSHEQH 115

oring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

tabase : PIR_68:*
          1: pir1:*
          2: pir2:*
          3: pir3:*
          4: pir4:*
          Listing first 45 summaries

arched: 219241 seqs, 76174552 residues

tal number of hits satisfying chosen parameters: 219241

nimum DB seq length: 0
ximum DB seq length: 2000000000

st-processing: Minimum Match 0%
               Maximum Match 100%

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	% Match	Query Length	DB	ID	Description
1	142	22.7	713	2	T44447	neuregulin-3 [impo
2	140	22.4	602	2	A45769	acetylcholine rece
3	126.5	20.2	462	2	I38404	neu differentiatio
4	126.5	20.2	640	2	A43273	neu differentiatio
5	124	19.8	637	2	C43273	heregulin precurso
6	123.5	19.8	639	2	C43273	heregulin precurso
7	123	19.7	645	2	B43273	neu differentiatio
8	122.5	19.6	162	2	S68401	heregulin, splice
9	119.5	19.1	1217	2	S68400	heregulin precurs
10	119	19.1	636	2	EGMSMG	epidermal growth f
11	117	18.7	636	2	I61718	neu differentiatio
12	116.5	18.6	125	2	S62676	heregulin isoform
13	116.5	18.6	1133	1	I38405	neu differentiatio
14	115	18.4	140	2	EGRT	epidermal growth f
15	115	18.4	140	2	B36837	DIL protein - vari
16	113.5	18.2	140	2	T28439	hypothetical prote
17	113	18.1	140	2	T30766	growth factor - va
18	112	17.9	53	2	S17294	epidermal growth f
19	112	17.9	140	2	C72150	B3R protein - vari
20	112	17.9	621	2	I38467	low density lipopr
21	112	17.9	850	2	JC5700	ErbB kinase activa
22	112	17.9	868	2	JC5702	ErbB kinase activa
23	109.5	17.5	175	2	JC5701	ErbB kinase activa
24	109.5	17.5	241	2	I38408	neu differentiatio
25	109.5	17.5	241	2	S33359	glial growth facto
26	109.5	17.5	296	2	D43273	heregulin precurso
27	109.5	17.5	296	2	A56943	sensory/motor neur
28	109	17.4	1207	1	S33257	glial growth facto
29	108.5	17.4	140	1	EGHU	epidermal growth f
29	108.5	17.4	140	1	WMVZ9	growth factor - va
30	107.5	17.2	230	2	A56210	neu differentiatio
31	107.5	17.2	662	2	I61722	neu differentiatio
32	105.5	16.9	294	2	A48844	TGF alpha-like pro
33	103.5	16.6	142	1	WMVZ3C	growth factor - va
34	103	16.5	85	1	EGVZM1	growth factor - my
35	.101	16.4	4543	1	A53102	alpha-2-macroglobu
36	100.5	16.1	80	2	EGVZSF	growth factor - ra
37	99.5	16.1	82	2	A61040	transforming growt
38	99.5	15.9	160	2	S39795	transforming growt
39	99.5	15.9	208	2	A38432	heparin-binding EG
40	99.5	15.9	208	1	A41914	diptheria toxin re
41	98.5	15.8	160	1	WFHU1	transforming growt
42	98.5	15.8	208	2	JC1410	heparin-binding EG
43	97.5	15.6	121	2	I58134	transforming growt
44	97.5	15.6	208	2	JC1409	heparin-binding EG
45	96	15.4	178	2	JC1467	betacellulin precu

ALIGNMENTS

RESULT 1
T44447
neuregulin-3 [imported] - mouse
C;Species: Mus musculus (house mouse)
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C;Accession: T44447
R;Zhang, D.; Sliwkowski, M.X.; Mark, M.; Frantz, G.; Akita, R.; Sun, Y.; Hillan, K.;
Proc. Natl. Acad. Sci. U.S.A. 94, 9562-9567, 1997
A;Title: Neuregulin-3 (NRG3): A novel neural tissue-enriched protein that binds and a
A;Reference number: Z22773; MUID:97420720
A;Status: Preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-713 <ZHA>
A;Cross-references: EMBL:AF010130; NID:g2429163; PIDN:AAB70914.1; PID:g2429164
A;Genetics:
A;Gene: NRG3

Query Match 22.7%; Score 142; DB 2; Length 713;
Best Local Similarity 26.6%; Pred. No. 6.1e-07;
Matches 38; Conservative 20; Mismatches 47; Indels 38; Gaps 6;

```
Qy    3 TDHEEPCGPSHKSPCLNGGLCYVIPTIPS--PFCRCVENYTGARCEEVFLP-------  51
          :::|| :||| ::||||||:||||| |||:     |::| |: ||||||:|
Db  286 SEHPKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQ-FLPKTDSILSDP 344

Qy   52 ---GSSIQTKSNLFE----------AFVALAVLVTLLIGAFYFLCRKGHFQRASSVQ---  95
           :::|          ||||  :  | |||  | ||||
Db  345 TDHLGIEFMESEDVYQRQVLSISCIIFGIVIVGMFCAAFVFKSKK----QAKIQEHLKE 400

Qy   96 -----YDINLVETSSTSAHHSH 112
             |  | ||||||:|   ||
Db  401 SQNGKNYSLKASSTKSESLMKSH 423
```

RESULT 2
A45769
acetylcholine receptor synthesis stimulator ARIA-1 precursor - chicken
C;Species: Gallus gallus (chicken)
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 21-Jul-2000
C;Accession: A45769
R;Falls, D.L.; Rosen, K.M.; Corfas, G.; Lane, W.S.; Fischbach, G.D.
Cell 72, 801-815, 1993
A;Title: ARIA, a protein that stimulates acetylcholine receptor synthesis, is a membe
A;Reference number: A45769; MUID:93201602
A;Accession: A45769
A;Molecule type: mRNA; protein
A;Residues: 1-602 <FAL>
A;Cross-references: GB:L11264; NID:g212603; PIDN:AAA49037.1; PID:g212604
A;Experimental source: brain

A;Note: sequence extracted from NCBI backbone (NCBIN:127787, NCBIP:127788)
A;Reference number:
A;Accession:
A;Title: Characterization of a breast cancer cell differentiation factor that speci
C;Superfamily: unassigned EGF-related proteins; EGF homology

J. Biol. Chem. 268, 18407-18410, 1993
A;Title: Characterization of a breast cancer cell differentiation factor that speci
A;Reference number: A48498; MUID:93366731
A;Accession: A48498
A;Molecule type: protein
A;Residues: 20-21,'X',23-24,'XX',27-28 <CUL>
R;Peles, E.; Bacus, S.S.; Koski, R.A.; Lu, H.S.; Wen, D.; Ogden, S.G.; Levy, R.B.; \
Cell 69, 205-216, 1992
A;Title: Isolation of the neu/HER-2 stimulatory ligand: a 44 kd glycoprotein that in
A;Reference number: A38155; MUID:92208945
A;Accession: A38155
A;Molecule type: protein
A;Residues: 'X',15-16,'X',18-20,'RG',23-24,'GP',27,'E',29,'XP',32-36 <PEL>
A;Note: sequence extracted from NCBI backbone (NCBIP:91347)
C;Genetics:
A;Gene: HGL
A;Cross-references: GDB:132656; OMIM:142445
A;Map position: 8p22-8p11
C;Superfamily: unassigned EGF-related proteins; EGF homology
F;182-221/Domain: EGF homology <EGF>

Query Match
Best Local Similarity   20.2%;   Score 126.5;   DB 2;   Length 640;
Matches   32;   Conservative   25;   Mismatches   46;   Indels   25;   Gaps   6;

Qy   3 TDHEEPCGPSHKSFCLNGGLCYIFPTIPSP---FCRCVENYTGARCEEVFLPGSSIQTKS 59
Db 176 TSHLVKCAEKKTFCVNGGECFMVKDLSNPSRYLCKCPQFGTGARCTE-NVP-MKVQNQE 233

Qy 60 NLFEAFVALAVLVT------LIIGAFYFLC------RKGHFQRASSVQYDINLVE 102
Db 234 KAEELYQKRVLTITGICIALLVVG---IMCVVAYCKTKKQRKKLHDRLRQSLRSERNNMM 290

Qy 103 TSSTSAHH 110
Db 291 NIANGPHH 298

RESULT   4
A43273
heregulin precursor, splice form alpha - human
N;Alternate names: breast cancer cell differentiation factor p45; Neu differentiation fa
A;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 11-Jan-2000
C;Accession: A43273; A48498; A38155
R;Holmes, W.E.; Sliwkowski, M.X.; Akita, R.W.; Henzel, W.J.; Lee, J.; Park, J.W.; Yansur
Science 256, 1205-1210, 1992
A;Title: Identification of heregulin, a specific activator of p185(erbB2).
A;Reference number: A43273; MUID:9227125
A;Accession: A43273
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-64 <HOL>
A;Experimental source: breast tumor cell line, MDA-MB-231, ATCC HTB 26
A;Note: sequence extracted from NCBI backbone (NCBIP:103250)
A;Molecule type: mRNA
R;Culouscou, J.M.; Plowman, G.D.; Carlton, G.W.; Green, J.M.; Shoyab, M.

RESULT   5
C43273
heregulin precursor, splice form beta-2 - human
A;Species: Homo sapiens (man)
Mol. Cell. Biol. 14, 1909-1919, 1994
A;Title: Structural and functional aspects of the multiplicity of Neu differentiatio
A;Reference number: A56210; MUID:94158863
A;Accession: I38407
A;Molecule type: mRNA
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Residues: 119-406 <RES>
A;Molecule type: mRNA
A;Cross-references: EMBL:U02329; NID:g408408; PIDN:AAA19954.1; PID:g408409
A;Reference number: A43273; MUID:9227125
A;Accession: C43273
A;Title: Identification of heregulin, a specific activator of p185(erbB2).
R;Holmes, W.E.; Sliwkowski, M.X.; Akita, R.W.; Henzel, W.J.; Lee, J.; Park, J.W.; Ya
Science 256, 1205-1210, 1992
C;Accession: C43273; I38407
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 17-Nov-2000
A;Reference number: I38407
A;Title: Structural and functional aspects of the multiplicity of Neu differentiatio
Mol. Cell. Biol. 14, 1909-1919, 1994
R;Wen, D.; Suggs, S.V.; Karunagaran, D.; Liu, N.; Cupples, R.L.; Luo, Y.; Janssen, A
A;Residues: 1-637 <HOL>
A;Molecule type: mRNA
A;Accession: C43273
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
A;Genetics:
A;Gene: GDB:HGL
A;Cross-references: GDB:132656; OMIM:142445
A;Map position: 8p22-8p11
C;Superfamily: unassigned EGF-related proteins; EGF homology
C;Keywords: alternative splicing
F;182-221/Domain: EGF homology <EGF>

Query Match   19.8%;   Score 124;   DB 2;   Length 637;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

January 8, 2002, 17:25:56 ; Search time 67.46 Seconds
(without alignments)
1168.312 Million cell updates/sec

nucleic - nucleic search, using sw model

n on:

rfect score: 348
quence: 1 atgccaacagatcacgaaga.........acagtcatgaacaacactga 348

tle: US-09-640-041-3

oring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

arched:

tal number of hits satisfying chosen parameters: 702406

351203 seqs, 113238999 residues

ximum DB seq length: 200000000
imum DB seq length: 0

st-processing: Minimum Match 0%
Maximum Match 100%
listing first 45 summaries

Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

-tabase :

SUMMARIES

esult No.	Score	% Query Match	Length	DB ID	Description
1	30.6	8.8	1559	4	US-09-019-095A-7	Sequence 7, Appl
2	29.8	8.6	2538	4	US-08-899-437-1	Sequence 1, Appli
3	29.8	8.6	2538	4	US-09-126-121-1	Sequence 1, Appl
4	29.6	8.5	5864	4	US-08-482-918-47	Sequence 47, Appl
5	29.6	8.5	5864	4	US-09-224-681-47	Sequence 47, Appl
6	29.6	8.5	5864	4	US-08-336-728A-47	Sequence 47, Appl
7	29.2	8.4	3807	4	US-08-482-918-43	Sequence 43, Appl
8	29.2	8.4	3807	4	US-09-224-681-43	Sequence 43, Appl
9	29.2	8.4	3807	4	US-08-336-728A-43	Sequence 43, Appl
10	29.2	8.4	5738	1	US-09-409-995-3	Sequence 3, Appl
11	29.2	8.4	5738	4	US-08-685-467-3	Sequence 3, Appl
12	29.2	8.4	7291	4	US-08-913-944-3	Sequence 3, Appl
13	28.4	8.2	2503	3	US-08-691-563C-60	Sequence 60, Appl
14	28	8.0	2282	1	US-08-373-579-5	Sequence 5, Appli
15	28	8.0	2282	2	US-08-418-595-5	Sequence 5, Appli
16	28	8.0	2282	2	US-08-665-926-5	Sequence 5, Appli
17	28	8.0	2282	4	US-09-162-437-5	Sequence 5, Appli
18	28	8.0	2282	4	US-08-740-240-5	Sequence 5, Appl
19	27.8	8.0	820	4	US-08-836-252A-5	Sequence 5, Appl
20	27.8	8.0	820	4	US-08-482-918-45	Sequence 45, Appl
21	27.8	8.0	820	4	US-09-224-681-45	Sequence 45, Appl
22	27.8	8.0	820	4	US-08-336-728A-45	Sequence 45, Appl
23	27.8	8.0	1873	4	US-09-019-095A-37	Sequence 37, Appl
24	27.4	7.9	1350	3	US-08-258-287B-41	Sequence 41, Appl
25	27.4	7.9	1350	3	US-08-368-704C-41	Sequence 41, Appl
26	27.4	7.9	1350	4	US-08-908-436-1	Sequence 1, Appl
27	27.2	7.8	1463	5	PCT-US93-05640-10	Sequence 10, Appl
28	27	7.8	2031	3	US-08-262-220-13	Sequence 13, Appl
29	27	7.8	2031	3	US-08-471-733-13	Sequence 13, Appl
30	27	7.8	2031	3	US-08-468-878-13	Sequence 13, Appl
31	27	7.8	2031	4	US-08-750-494-13	Sequence 13, Appl
32	27	7.8	2075	3	US-08-262-220-3	Sequence 3, Appli
33	27	7.8	2075	3	US-08-471-733-3	Sequence 3, Appli
34	27	7.8	2075	3	US-08-468-878-3	Sequence 3, Appli
35	26.6	7.6	1425	4	US-08-750-494-3	Sequence 3, Appli
36	26.6	7.6	1425	3	US-09-729-416C-10	Sequence 10, Appl
37	26.6	7.6	1785	3	US-08-729-416C-16	Sequence 16, Appl
38	26.4	7.6	1200	4	US-09-327-681-5	Sequence 5, Appli
39	26.4	7.6	1708	4	US-08-991-677-1	Sequence 1, Appli
40	26.4	7.6	1757	4	US-08-199-779-2	Sequence 2, Appli
41	26.4	7.6	1757	3	US-08-578-615A-105	Sequence 105, App
42	26.4	7.6	1757	3	PCT-US94-07770-105	Sequence 105, App
43	26.4	7.6	1812	1	US-08-199-779-1	Sequence 1, Appli
44	26.4	7.6	1812	3	US-08-578-615A-104	Sequence 104, App
45	26.4	7.6	1812	5	PCT-US94-07770-104	Sequence 104, App

ALIGNMENTS

RESULT 1
US-09-019-095A-7
Sequence 7, Application US/09019095A
Patent No. 6287858.
GENERAL INFORMATION:
APPLICANT: D'Andrea, Alan D.
APPLICANT: Zhu, Yuan
TITLE OF INVENTION: Deubiquitinating Enzymes That Regulate
TITLE OF INVENTION: Cell Growth
FILE REFERENCE: DFCI-435p2A2
CURRENT APPLICATION NUMBER: US/09/019,095A
CURRENT FILING DATE: 1998-02-05
PRIOR APPLICATION NUMBER: PCT/US96/12884
PRIOR FILING DATE: 1996-08-07
PRIOR APPLICATION NUMBER: US 60/002,066
PRIOR FILING DATE: 1995-08-09
PRIOR APPLICATION NUMBER: US/ 60/019,787
PRIOR FILING DATE: 1996-06-14
NUMBER OF SEQ ID NOS: 51
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 7
LENGTH: 1559
TYPE: DNA
ORGANISM: Murine
US-09-019-095A-7

Query Match   8.8%;   Score 30.6; DB 4; Length 1559;
Best Local Similarity   60.0%;   Pred. No. 0.53;
Matches 51; Conservative   0; Mismatches 34; Indels   0; Gaps   0;

Qy   137 aggaggtttttctcccaggctccaagcatccaaactaaaagtaacctgtttgaagctttg 196
           ||||| |||| ||| || || ||| ||||||
Db   1322 aagagatgacattctgagctgtcttgcagaaccacacaccaacttgtgaggaggcttgtc 1381

Qy   197 tggcattggcggtcctagtaacact 221
           ||| ||||| ||||| | || |||
Db   1382 tgggattggctgtcctgggaaagact 1406

RESULT 2
US-08-899-437-1
Sequence 1, Application US/08899437
Patent No. 6121415.
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J., Mark, Melanie Rose; Zhang, Dong Xiao
TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
TITLE OF INVENTION: Ligands and Uses Therefor
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:

```
C
R   EMBL; AF083067; AAD21874.1; -.
R   InterPro; IPR000561; EGF-like.
R   InterPro; IPR001336; EGF_1.
R   Pfam; PF00008; EGF.
R   PRINTS; PR00009; EGFGF.
R   SMART; SM00181; EGF; 1.
R   PROSITE; PS00022; EGF_1; 1.
R   PROSITE; PS01186; EGF_2; FALSE_NEG.
W   Growth factor; EGF-like domain; Glycoprotein; Transmembrane;
W   Multigene family; Alternative splicing.
T   CHAIN         1    115   PRO-NEUREGULIN-4.
T   CHAIN         1     61   NEUREGULIN-4.
T   CHAIN        62    115   MEMBRANE-BOUND FORM.
T   TRANSMEM     63     83   INTERNAL SIGNAL SEQUENCE (POTENTIAL).
T   DOMAIN       84    115   CYTOPLASMIC (POTENTIAL).
T   DOMAIN        5     46   EXTRACELLULAR (POTENTIAL).
T   DOMAIN        9     23   EGF-LIKE.
T   DOMAIN       23     34   BY SIMILARITY.
T   DISULFID     17     27   BY SIMILARITY.
T   DISULFID     36     45   BY SIMILARITY.
T   DISULFID     39     39   BY SIMILARITY.
T   CARBOHYD     36          N-LINKED (GLCNAC...) (POTENTIAL).
T   CARBOHYD     39          N-LINKED (GLCNAC...) (POTENTIAL).
T   CARBOHYD     60          (POTENTIAL).
T   SEQUENCE    115 AA;  12743 MW;  989A1E3376F857B49 CRC64;

Query Match        78.2%;  Score 489;  DB 1;  Length 115;
Best Local Similarity   75.7%;  Pred. No. 2.8e-45;
Matches   87;  Conservative  12;  Mismatches  16;  Indels  0;  Gaps  0;

Qy    1 MPTDHEEPCGPSHKSFCLNGGLCYVIPTIPSPFCRCVENYTGARCEEVFLPGGSSIQTKSN 60
         ||||||:|||| |:|||||||:||||||||||||||:||||||||||||||: :||: :||
Db    1 MPTDHEQPCGPRHRSFCLNGGICYVIPTIPSPFCRCIENYTGARCEEVFLPSSSIPSESN 60

Qy   61 LFEAFVALAVLVTLLIIGAFYFLCRKGHFQRASSVOYDINLVETSSTSAHHSHEOH 115
         | ||| | |||| |  | | :|||||||| ||||||   ||   |:|| | || :|
Db   61 LSAAFVVLAVLLTLTIAALCFLCRKGHLQRASSVOCEISLVETNNTRTRHSHREH 115

RESULT 2
NRG3_MOUSE
ID   NRG3_MOUSE      STANDARD;      PRT;   713 AA.
AC   O35181;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PRO-NEUREGULIN-3 PRECURSOR (PRO-NRG3) [CONTAINS: NEUREGULIN-3 (NRG-
DE   3)].
GN   NRG3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97420720; PubMed=9275162;
RA   Zhang D., Sliwkowski M.X., Mark M., Frantz G., Akita R., Sun Y.,
RA   Hillan K., Crowley C., Brush J., Godowski P.J.;
RT   "Neuregulin-3 (NRG3): a novel neural tissue-enriched protein that
RT   binds and activates ErbB4.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:9562-9567(1997).
-!- FUNCTION: DIRECT LIGAND FOR THE ERBB4 TYROSINE KINASE RECEPTOR.
    BINDING RESULTS IN LIGAND-STIMULATED TYROSINE PHOSPHORYLATION AND
    ACTIVATION OF THE RECEPTOR. DOES NOT BIND TO THE EGF RECEPTOR,
    ERBB2 OR ERBB3 RECEPTORS.
-!- SUBCELLULAR LOCATION: EXISTS AS AN TYPE I MEMBRANE PROTEIN AND AS
    A PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE
    MEMBRANE-BOUND FORM DOES NOT SEEM TO BE ACTIVE (BY SIMILARITY).
-!- TISSUE SPECIFICITY: EXPRESSED IN SYMPATHETIC, MOTOR, AND SENSORY
    NEURONS.
-!- DEVELOPMENTAL STAGE: DETECTED AS EARLY AS E11. IN E13 EMBRYOS,
    DETECTED MAINLY IN THE NERVOUS SYSTEM. IN E16 EMBRYOS, DETECTED IN
    THE BRAIN, SPINAL CORD, TRIGEMINAL, VESTIBULAR-COCHLEAR, AND
```

```
CC   SPINAL GANGLIA, IN ADULTS, EXPRESSED IN SPINAL CORD, AND NUMEROUS
CC   BRAIN REGIONS.
-!- DOMAIN: THE CYTOPLASMIC DOMAIN MAY BE INVOLVED IN THE REGULATION
    OF TRAFFICKING AND PROTEOLYTIC PROCESSING. REGULATION OF THE
    PROTEOLYTIC PROCESSING INVOLVES INITIAL INTRACELLULAR DOMAIN
    DIMERIZATION (BY SIMILARITY).
-!- DOMAIN: ERBB RECEPTOR BINDING IS ELICITED ENTIRELY BY THE EGF-LIKE
    DOMAIN (BY SIMILARITY).
-!- PTM: PROTEOLYTIC CLEAVAGE CLOSE TO THE PLASMA MEMBRANE ON THE
    EXTERNAL FACE LEADS TO THE RELEASE OF THE SOLUBLE GROWTH FACTOR
    FORM (BY SIMILARITY).
-!- PTM: EXTENSIVE GLYCOSYLATION PRECEDES THE PROTEOLYTIC CLEAVAGE (BY
    SIMILARITY).
-!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
-!- SIMILARITY: BELONGS TO THE NEUREGULIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; AF010130; AAA70914.1; -.
MGD; MGI:1097165; Nrg3.
InterPro; IPR000561; EGF-like.
InterPro; IPR002154; Neuregulin.
Pfam; PF00008; EGF.
Pfam; PF02158; Neuregulin; 1.
SMART; SM00181; EGF; 1.
PROSITE; PS00022; EGF_1; 1.
PROSITE; PS01186; EGF_2; 1.
Growth factor; EGF-like domain; Transmembrane; Multigene family.
CHAIN          1    713   PRO-NEUREGULIN-3.
CHAIN          1    361   NEUREGULIN-3.
CHAIN        362    713   MEMBRANE-BOUND FORM.
TRANSMEM     363    383   INTERNAL SIGNAL SEQUENCE (POTENTIAL).
DOMAIN       384    713   CYTOPLASMIC (POTENTIAL).
DOMAIN         1    362   EXTRACELLULAR (POTENTIAL).
DOMAIN       105    288   SER/THR-RICH.
DOMAIN       287    331   EGF-LIKE.
DOMAIN        13     21   POLY-ALA.
DOMAIN        26     34   POLY-ALA.
DOMAIN       127    135   POLY-THR.
DOMAIN       250    253   POLY-ALA.
DOMAIN       254    263   POLY-ALA.
DOMAIN       264    267   POLY-THR.
DISULFID     292    306   BY SIMILARITY.
DISULFID     300    319   BY SIMILARITY.
DISULFID     321    330   BY SIMILARITY.
SEQUENCE     713 AA;  77369 MW;  9F7D1D5E7FC8DCF0 CRC64;

Query Match        22.7%;  Score 142;  DB 1;  Length 713;
Best Local Similarity   26.6%;  Pred. No. 1.4e-07;
Matches   38;  Conservative  20;  Mismatches  47;  Indels  38;  Gaps  6;

Qy    3 TDHEEPCGPSHKSFCLNGGLCYVIPTIPS--PFCRCVENYTGARCEEVFLP--------- 51
         ||:|||   ||:|||||||||||||:|| |   |||  ||| |||||:||||
Db    3 SEHFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQ-FLPKIDSILSDP 344

Qy   52 ----GSSIQTKSNLFE-------AFVALAVLVTLLIIGAFYFLCRKGHFQRASSVO----- 95
            |:   |:||:||:::     :||||||| |   |||:||||:|||||:|||:|
Db  286 SEIFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQ-FLPKIDSILSDP 344

Qy   96 ------YDINLVETSSTSAHHSH 112
              ||:|  |:|| | || :|
Db  345 TDHLGIEFMESEDYQROVLSICIIFGIVIVGMFCAAFYFKSKK-----QAKQIQEHLKE 400

Db  401 SQNGKNYSLKASSTKSESLMKSH 423
```

```
RESULT 3
NRG1_CHICK
ID   NRG1_CHICK      STANDARD;      PRT;   602 AA.
```

```
                              GenCore version 4.5
                              Copyright (c) 1993 - 2000 Compugen Ltd.

nucleic - nucleic search, using sw model

on on:        January 8, 2002, 17:27:15 ; Search time 158.12 Seconds
                              (without alignments)
                              1886.853 Million cell updates/sec

tle:          US-09-640-041-3
rfect score:  348
quence:       1 atgccaacagatcacgaaga.........acagtcatgaacaacactga 348

oring table:  IDENTITY_NUC
              Gapop 10.0 , Gapext 1.0
arched:       930621 seqs, 428662619 residues

tal number of hits satisfying chosen parameters:    1861242

nimum DB seq length: 0
ximum DB seq length: 200000000

st-processing: Minimum Match 0%
               Maximum Match 100%
               Listing first 45 summaries

tabase :       N_Geneseq_1101:*
          1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
          2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
          3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
          4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
          5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
          6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
          7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
          8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
          9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
         10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
         11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
         12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
         13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
         14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
         15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
         16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
         17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
         18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
         19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
         20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
         21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
         22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*
         22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	% Query Match	Length	DB	ID	Description
1	348	100.0	348	22	AAA91294	Human EGFH2 coding
2	252	72.4	348	22	AAA91295	Mouse EGFH2 coding
3	32	9.2	18	18	AAT72320	Embryonic stem cel
4	31.6	9.1	363	22	AAF64385	Novel human polynu
5	30.6	8.8	1559	18	AAF91855	DUB-1 enhancer/pro
6	30.6	8.8	92407	22	AAF28549	Genomic fragment #
7	30.4	8.7	803	22	AAS22969	DNA encoding novel
8	30.4	8.7	4487	7	AAN60943	Sequence encoding
9	30.2	8.7	783	21	AAC33642	Arabidopsis thalia
10	29.8	8.6	1202	22	AAH46934	Human secreted pro
11	29.8	8.6	1947	22	AAH46950	Human secreted pro

```
12  29.8  8.6  2538  20  AAX06987   Mouse neuregulin r
13  29.6  8.5  5864  22  AAS10458   Human stem cell fa
14  29.6  8.5  5864  22  AAH41342   Human stem cell fa
15  29.6  8.5  5864  22  AAS04122   Human SCF (stem ce
16  29.6  8.5  5864  22  AAS04223   Human SCF (stem ce
17  29.6  8.5  5864  22  AAS04223   Human SCF (stem ce
18  29.6  8.5  5864  22  AAH23900   Human stem cell fa
19  29.6  8.5  5864  22  AAP89102   Human stem cell fa
20  29.4  8.4  462  22  AAC56228   Pinus radiata tran
21  29.4  8.4  564  22  AAI39710   Probe #8396 used t
22  29.4  8.4  352  21  AAA13769   Human stem cell fa
23  29.2  8.4  353  16  AAT04888   Human SCF genomic
24  29.2  8.4  863  21  AAT04888   Barley sucrose pho
25  29.2  8.4  1215  21  AAZ34773   Arabidopsis thalia
26  29.2  8.4  3807  22  AAH41340   Human genomic SCF
27  29.2  8.4  3807  22  AAC44993   Human genomic SCF
28  29.2  8.4  3807  22  AAS04120   Human genomic SCF
29  29.2  8.4  3807  22  AAS10456   Human stem cell f
30  29.2  8.4  3807  22  AAS04221   Human stem cell fa
31  29.2  8.4  3807  22  AAH23898   Human genomic SCF
32  29.2  8.4  4063  22  AAF89100   Human stem cell fa
33  29.2  8.4  4063  22  AAQ11540   Human stem cell Fa
34  29    8.4  7253  22  AAQ11540   Haemophilus influe
35  29    8.3  7291  17  AAA92499   Haemophilus influe
36  29    8.3  376  18  AAT41476   Staphylococcus aur
37  29    8.3  523  21  AAV78333   Arabidopsis thalia
38  29    8.3  753  21  AAC38468   Arabidopsis thalia
39  28.8  8.3  760  21  AAC35184   Arabidopsis thalia
40  28.8  8.3  963  22  AAC51933   Human olfactory re
41  28.8  8.3  308  21  AAH31919   Human olfactory re
42  28.8  8.3  822  20  AAC03591   Human secreted pro
43  28.8  8.3  1389  20  AAH48641   Human ceramidase K
44  28.6  8.2  1430  21  AAZ17120   Human gene express
45  28.6  8.2  376  21  AAZ24191   Murine PG-Lb cDNA.
     28.6  8.2  7980  19  AAV44243   Lettuce resistance
     28.6  8.2  975  21  AAH31732   Human olfactory re
     28.6  8.2  1260  22  AAH99677   Human protein enco
```

RESULT   1
AAA91294
ID   AAA91294 standard; DNA; 348 BP.
XX
AC   AAA91294;
XX
DT   19-JUN-2001 (first entry)
XX
DE   Human EGFH2 coding sequence.
XX
KW   EGFH2; epidermal growth factor H2; EGFH2 protein modulated disorder;
KW   cancer; peripheral neuropathy; amyotrophic lateral sclerosis; therapy;
KW   Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW   ischaemic stroke; brain injury; acute spinal cord injury; infection;
KW   nervous system injury; multiple sclerosis; dementia; epilepsy;
KW   peripheral nerve injury; acoustic trauma; human; ds.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..348
FT   /*tag= a
FT   /product= "EGFH2"
XX
PN   WO20011A415-A2.
XX
PD   01-MAR-2001.
XX
PF   15-AUG-2000; 2000WO-US22326.
XX
PR   20-AUG-1999; 99US-0149986.
XX
PA   (CHIR) CHIRON CORP.
XX
```

ALIGNMENTS

```
CC  -!- DOMAIN: ERBB RECEPTOR BINDING IS ELICITED ENTIRELY BY THE EGF-LIKE
CC      DOMAIN.
CC  -!- PTM: PROTEOLYTIC CLEAVAGE CLOSE TO THE RELEASE OF THE SOLUBLE GROWTH FACTOR
CC      EXTERNAL FACE LEADS TO THE RELEASE OF THE SOLUBLE GROWTH FACTOR
CC      FORM (BY SIMILARITY).
CC  -!- PTM: EXTENSIVE GLYCOSYLATION PRECEDES PROTEOLYTIC CLEAVAGE (BY
CC      SIMILARITY).
CC  -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC  -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC  -!- SIMILARITY: BELONGS TO THE NEUREGULIN FAMILY.
CC
DR  EMBL; U96612; AAB71812.1; -.
DR  HSSP; Q12784; 1HRE.
DR  InterPro; IPR000561; EGF-like.
DR  InterPro; IPR003598; Ig_C2.
DR  InterPro; IPR003006; Ig_MHC.
DR  InterPro; IPR002154; Neuregulin.
DR  Pfam; PF00008; EGF; 1.
DR  Pfam; PF00047; Ig; 1.
DR  Pfam; PF02158; Neuregulin; 1.
DR  PRINTS; PR01089; NEUREGULIN.
DR  SMART; SM00181; EGF; 1.
DR  SMART; SM00408; IGC2; 1.
DR  PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR  PROSITE; PS01186; EGF_2; 1.
KW  Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;
KW  Transmembrane; Alternative splicing.
FT  PROPEP      1     13       BY SIMILARITY.
FT  CHAIN      14    461       PRO-NEUREGULIN-1.
FT  CHAIN      14    241       NEUREGULIN-1, MEMBRANE-BOUND FORM.
FT  DOMAIN     14    242       BY SIMILARITY.
FT  TRANSMEM  243    265       INTERNAL SIGNAL SEQUENCE (POTENTIAL).
FT  DOMAIN    266    461       EXTRACELLULAR (POTENTIAL).
FT  DOMAIN     50    119       IG-LIKE C2-TYPE DOMAIN.
FT  DOMAIN    165    177       SER/THR-RICH.
FT  DOMAIN    178    222       EGF-LIKE.
FT  DISULFID   57    112       BY SIMILARITY.
FT  DISULFID  182    196       BY SIMILARITY.
FT  DISULFID  190    210       BY SIMILARITY.
FT  DISULFID  212    221       BY SIMILARITY.
FT  CARBOHYD   73     73       N-LINKED (GLCNAC...)     (POTENTIAL).
FT  CARBOHYD  120    120       N-LINKED (GLCNAC...)     (POTENTIAL).
FT  CARBOHYD  126    126       N-LINKED (GLCNAC...)     (POTENTIAL).
FT  CARBOHYD  164    164       N-LINKED (GLCNAC...)     (POTENTIAL).
SQ  SEQUENCE  461 AA;  50890 MW;  935C9560F7148336 CRC64;

RESULT   2
ID   Q9BDQ0
AC   Q9BDQ0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EPIDERMAL GROWTH FACTOR PRECURSOR.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

Query Match   20.9%; Score 130.5; DB 11; Length 461;
Best Local Similarity 25.8%; Pred. No. 1.4e-06;
Matches 33; Conservative 25; Mismatches 45; Indels 25; Gaps 6;

```
Qy      3 TDHEEPCGPSHKSFCLNGGLCYVIPTIPSP---FCRCVENYTGARCEEVFLPGSSIQTKS 59
          |  |  ::||:: |:|:|||| |:|  : ||   ||||:|:||| |::|:||| :| ::|:
Db    176 TSHLVKCAEKKTFCVNGGECFMVKDLSNPSRYLCKCQPGFTGARCTE-NVP-MKVQTQE 233

Qy     60 NLFEAFVALAVLVT-----LIIGAFYFLC--------RKGHFQRASSVQYDINLVE 102
          ||::||:||:|:     :|||||:|||        |:||:| |:||| :: | |:
Db    234 KAEELYQKRVLTITGICIALLVVG---IMCVVAYCKTKKQRQKLHDRLLQSLRSERNNV 290
```

Qy   103 TSSTSAHH 110
          || ||
Db   291 NIANGPHH 298

```
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kim J.G., Vallet J.L., Christenson R.K.;
RT   "Characterization of uterine epidermal growth factor during pregnancy
RT   in pigs.";
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF336151; AAK18830.1; -.
DR   CHAIN      970   1022       EPIDERMAL GROWTH FACTOR.
SQ   SEQUENCE  1214 AA;  133505 MW;  83D104090574 58CF CRC64;
```

Query Match   20.6%; Score 128.5; DB 6; Length 1214;
Best Local Similarity 33.7%; Pred. No. 6.2e-06;
Matches 29; Conservative 14; Mismatches 34; Indels 9; Gaps 2;

```
Qy      9 CGPSHKSFCLNGGLCYVIPTIPSPFCRCVENYTGARCEEVFLP-----GSSIQTKSNLFE 63
          ||::||:|| | : |  ||| ||:||||:|:||| |::|:|||     |::|:|| |:
Db    975 CPPSHDGYCLHGGVCMYIEAVDSYACNCVFGYVGERCQHRDLKWWELIRHAGLGRQWNVTV 1034

Qy     64 AFVALAVLVTLII-----GAFYFLCRK 85
          ||:||:|: | :      ||:|||||:
Db   1035 VAVCVVVLLLLLGLWGAHYRTQK 1060
```

```
RESULT   3
ID   Q9ESA9
AC   Q9ESA9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SMDF NEUREGULIN ALPHA 2B (FRAGMENT).
GN   NRG1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BDIX;
RA   Carroll S.L., Anderson K.D., Frohnert P.W.;
RT   "Structural and Functional Diversity of SMDF Neuregulin Splice
RT   Variants Expressed in the Adult Rat Nervous System.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF194440; AAG28429.1; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR002154; Neuregulin.
DR   InterPro; IPR002114; PTS_HPr_ser.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF02158; Neuregulin; 1.
DR   PRINTS; PR01089; NEUREGULIN.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS00589; PTS_HPR_SER; UNKNOWN_1.
FT   NON_TER     1      1
FT   NON_TER   298    298
SQ   SEQUENCE  298 AA;  32851 MW;  BD76F014C2B33026 CRC64;
```

Query Match   19.8%; Score 123.5; DB 11; Length 298;
Best Local Similarity 25.0%; Pred. No. 5.3e-06;
Matches 32; Conservative 25; Mismatches 46; Indels 25; Gaps 6;

```
Qy      3 TDHEEPCGPSHKSFCLNGGLCYVIPTIPSP---FCRCVENYTGARCEEVFLPGSSIQTKS 59
          |  |  ::||:: |:|:|||| |:|  : ||   ||||:|:||| |::|:||| :| ::|:
Db     46 TSHLIKCAEKEKTFCVNGGECFTVKDLSNPSRYLCKCQPGFTGARCTE-NVP-MKVQTQE 103

Qy     60 NLFEAFVALAVLVT-----LIIGAFYFLC--------RKGHFQRASSVQYDINLVE 102
          ||::||:||:|:     :|||||:|||        |:||:| |:||| :: | |:
Db    104 KAEELYQKRVLTITGICIALLVVG---IMCVVAYCKTKKQRQKLHDRLLQSLRSERNLV 160
```

```
                    GenCore version 4.5
                    Copyright (c) 1993 - 2000 Compugen Ltd.

1 nucleic - nucleic search, using sw model

Run on:     January 8, 2002, 14:00:50 ; Search time 1511.43 Seconds
                                         (without alignments)
                                         2474.167 Million cell updates/sec

title:
perfect score:  348
Sequence:  1 atgccaacagatcacgaaga..........acagtcatgaacaacactga 348

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:   1351937 seqs, 537289281 residues

Total number of hits satisfying chosen parameters:   22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   EST:*

   1: em_estfun:*
   2: em_esthum:*
   3: em_estin:*
   4: em_estom:*
   5: em_estpl:*
   6: em_estba:*
   7: em_estro:*
   8: em_estov:*
   9: em_htc:*
  10: gb_est1:*
  11: gb_est2:*
  12: gb_htc:*
  13: gb_gss:*
  14: em_gss_fun:*
  15: em_gss_hum:*
  16: em_gss_inv:*
  17: em_gss_pln:*
  18: em_gss_pro:*
  19: em_gss_rod:*
  20: em_gss_vrt:*
  21: em_gss_other:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	266	76.4	788	11	BF126566	BF126566 601650820
2	252	72.4	568	10	AA238077	AA238077 mx10e03.r
3	246.4	70.8	545	11	BF061527	BF061527 7j52f08.x
4	246.4	70.8	595	10	BE787057	BE787057 601476583
5	244.8	70.3	513	10	AI743118	AI743118 wg87d12.x
6	213	61.2	1149	11	BF793585	BF793585 60225 4235
7	194.4	55.9	374	11	H56480	H56480 yt87e09.s1
8	172.2	49.5	656	10	AA763377	AA763377 vw53f02.r
9	171.4	49.3	461	10	BF564092	BF564092 UI-R-C4-a
10	171.4	49.3	461	10	AW530086	AW530086 UI-R-C4-a
11	167	48.0	635	10	BG083489	BG083489 H309G04-
12	166	47.7	597	11	BE910804	BE910804 601661856

	No.	Score	%	Length	DB	ID	Description
c	13	115.2	33.1	496	10	BE627914	BE627914 un25a04.x
	14	100.2	28.8	520	10	AW106678	AW106678 um32b11.y
	15	81.8	23.5	240	10	BB569392	BB569392
	16	81.8	23.3	240	10	BB569392	BB569392
	17	81	23.3	243	10	AU059620	AU059620
	18	81	17.5	300	11	C94578	C94578 C94578 P1g
	19	61	17.5	547	10	BE624667	BE624667 un25a04.y
	20	45	12.9	385	10	AA960000	AA960000 vw53f02.s
c	21	35.4	10.2	382	10	AA960000	AA960000 341541 BA
	22	35.2	10.1	584	10	BH016371	BH016371 TDCL34TH
	23	34.8	9.9	1009	10	AL547748	AL547748 AL547748
	24	34.6	9.9	236	10	BE479108	BE479108 163676 BA
	25	34.4	9.9	557	13	AQ591571	AQ591571 HS_5320_A
c	26	34	9.8	570	13	AQ498222	AQ498222 HS_5104_A
	27	33.8	9.7	494	10	AV569930	AV569930 AV569930
c	28	33.6	9.7	431	10	AV556329	AV556329 AV556329
	29	33.6	9.5	879	11	BG178252	BG178252 60329984
	30	33.2	9.5	299	13	B71998	B71998 RPCI11-7D20
c	31	33.2	9.5	890	11	BF181171	BF181171 60180S834
	32	33	9.5	531	13	AQ179732	AQ179732 HS_3195_A
c	33	33	9.5	658	11	BG301355	BG301355 kt04d10.y
c	34	32.8	9.5	878	11	BE620125	BE620125 60148308
c	35	32.8	9.4	439	11	BG711530	BG711530 pg11n.pk0
	36	32.6	9.4	300	10	AI630255	AI630255 ad07e0i.x
c	37	32.6	9.4	426	11	AI631608	AI631608 wb01a06.x
	38	32.6	9.4	447	13	AQ430053	AQ430053 HS_5061_B
	39	32.6	9.4	478	11	BG159836	BG159836 de57f01.x
	40	32.6	9.4	587	10	AW549138	AW549138 L0046F07-
	41	32.6	9.4	641	11	BG073298	BG073298 H3121A03-
c	42	32.4	9.3	796	10	A2706084	A2706084 RPCI-23-2
	43	32.2	9.3	1093	13	AL311955	AL311955 Tetraodon
	44	32.2	9.2	522	10	AV540630	AV540630
	45	32	9.2	527	11	BF015717	BF015717 uy28b02.y
		32	9.2	592	13	CNS06YHZ	CNS06YHZ
```

ALIGNMENTS

RESULT 1
BF126566
```
LOCUS       BF126566       788 bp    mRNA       EST       24-OCT-2000
DEFINITION  601650820F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4071605 5',
            mRNA sequence.
ACCESSION   BF126566
VERSION     BF126566.1  GI:10965606
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 788)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM917 row: h column: 06
            High quality sequence.stop: 518.
FEATURES             Location/Qualifiers
     source          1..788
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NIH_MGC_76"
                     /clone="IMAGE:4071605"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
```

Column 1 (RESULT 1, continued)

```
PT   Novel epidermal growth factor H2 (EGFH2) polypeptides and
PT   polynucleotides encoding them useful for treating EGFH2
PT   protein-modulated disorders such as cancers, Alzheimer's disease,
PT   ischaemic stroke and epilepsy -
XX
PT   Claim 10; Fig 1; 48pp; English.
XX
CC   This sequence is the human epidermal growth factor H2 (EGFH2)
CC   protein of the invention. The EGFH2 DNA and protein sequences are useful
CC   for modulating the amount of EGFH2 protein in a subject, or treating an
CC   EGFH2 protein modulated disorder. The DNA or its antisense construct,
CC   ribozyme or a retroviral vector comprising a promoter are also useful
CC   for treating breast, prostate, pancreatic, oral or ovarian cancers,
CC   peripheral neuropathy, amyotrophic lateral sclerosis, Alzheimer's
CC   disease, Parkinson's disease, Huntington's disease, ischaemic stroke,
CC   brain injury, acute spinal cord injury, nervous system injury, multiple
CC   sclerosis, infection, dementia, epilepsy, peripheral nerve injury,
CC   acoustic trauma and tissue wound. The DNA and protein are also useful for
CC   providing trophic support for cells in a patient by implanting cells
CC   expressing them encapsulated in a semipermeable membrane. EGFH2
CC   polypeptides can be used to screen peptide libraries. Antibodies which
CC   bind to EGFH2 and/or variant polypeptides can be used in diagnosing and
CC   determining the prognosis of cancer, tumour progression,
CC   hyperproliferative cell growth or accompanying biological and physical
CC   manifestations.
XX
SQ   Sequence   115 AA;

Query Match            100.0%;  Score 625;  DB 22;   Length 115;
Best Local Similarity  100.0%;  Pred. No. 1.9e-51;
Matches  115;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 MPTDHEEPCGPSHKSFCLNGGLCVYIPTIPSPFCRCVENYTGARCEEVFLPGSSIQTKSN 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 mptdheepcgpshksfclnggicvyiptipspfcrcvenytgarceevflpgssiqtksn 60

Qy   61 LFEAFVALAVLVTLIIGAFYFLCRKGHFQRASSVQYDINLVETSSTSAHHSHEQH 115
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 lfeafvalavlvtliigafyflcrkghfqrassvqydinlvetsstsahhsheqh 115
```

RESULT 2

```
AC   AAY97698
ID   AAY97698 standard; Protein; 115 AA.
XX
AC   AAY97698;
XX
DT   19-JUN-2001  (first entry)
XX
DE   Mouse EGFH2 protein sequence.
XX
PN   WO200114415-A2.
XX
PD   01-MAR-2001.
XX
PF   15-AUG-2000; 2000WO-US22326.
XX
PR   20-AUG-1999; 99US-0149986.
XX
PA   (CHIRON ) CHIRON CORP.
PI   Kavanaugh WM, Cen H, Lee P;
XX
DE   EGFH2; epidermal growth factor H2; EGFH2 protein modulated disorder;
KW   cancer; peripheral neuropathy; amyotrophic lateral sclerosis; therapy;
KW   Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW   ischaemic stroke; brain injury; acute spinal cord injury; infection;
KW   nervous system injury; multiple sclerosis; dementia; epilepsy;
KW   peripheral nerve injury; acoustic trauma; mouse.
OS   mus sp.
```

Column 2 (RESULT 2, continued)

```
DR   WPI; 2001-211304/21.
DR   N-PSDB; AAA91295.
XX
PT   Novel epidermal growth factor H2 (EGFH2) polypeptides and
PT   polynucleotides encoding them useful for treating EGFH2
PT   protein-modulated disorders such as cancers, Alzheimer's disease,
PT   ischaemic stroke and epilepsy -
XX
PT   Claim 10; Fig 1; 48pp; English.
XX
CC   This sequence is the mouse epidermal growth factor H2 (EGFH2)
CC   protein of the invention. The EGFH2 DNA and protein sequences are useful
CC   for modulating the amount of EGFH2 protein in a subject, or treating an
CC   EGFH2 protein modulated disorder. The DNA or its antisense construct,
CC   ribozyme or a retroviral vector comprising a promoter are also useful
CC   for treating breast, prostate, pancreatic, oral or ovarian cancers,
CC   peripheral neuropathy, amyotrophic lateral sclerosis, Alzheimer's
CC   disease, Parkinson's disease, Huntington's disease, ischaemic stroke,
CC   brain injury, acute spinal cord injury, nervous system injury, multiple
CC   sclerosis, infection, dementia, epilepsy, peripheral nerve injury,
CC   acoustic trauma and tissue wound. The DNA and protein are also useful for
CC   providing trophic support for cells in a patient by implanting cells
CC   expressing them encapsulated in a semipermeable membrane. EGFH2
CC   polypeptides can be used to screen peptide libraries. Antibodies which
CC   bind to EGFH2 and/or variant polypeptides can be used in diagnosing and
CC   determining the prognosis of cancer, tumour progression,
CC   hyperproliferative cell growth or accompanying biological and physical
CC   manifestations.
XX
SQ   Sequence   115 AA;

Qy    1 MPTDHEEPCGPSHKSFCLNGGLCVYIPTIPSPFCRCVENYTGARCEEVFLPGSSIQTKSN 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 mptdheepcgpshksfclnggicvyiptipspfcrcvenytgarceevflpgssiqtksn 60

Qy   61 LFEAFVALAVLVTLIIGAFYFLCRKGHFQRASSVQYDINLVETSSTSAHHSHEQH 115
         |||:||||||:|||||||:|||||||:|||||||||||||||||||||||||||
Db   61 lfeatvalavivtliligatyflcrkghgqrassvqydinlvetsstsahhsheqh 115
```

RESULT 3

```
RESULT   3
AAW97617
ID   AAW97617 standard; Protein; 713 AA.
XX
AC   AAW97617;
XX
DT   10-MAY-1999  (first entry)
XX
DE   Mouse neuregulin related ligand NRG3.
XX
KW   Neuregulin related ligand; NRG3; mouse; ErbB4 receptor;
KW   signal transduction; nervous system disorder; neurodegeneration;
KW   neuropathy; therapy; diagnosis.
OS   Mus sp.
XX
FH   Key               Location/Qualifiers
FH
FT   Domain            1..362
FT                     /note= "extracellular domain, specifically claimed
FT                     in Claim 5(a)"
FT   Region            66..91
FT                     /note= "hydrophobic region"
FT   Region            105..286
FT                     /note= "mucin-like Ser/Thr-rich region, contains
FT                     sites for O-linked glycosylation"
FT   Domain            287..334
FT                     /note=
FT   Domain            363..385
FT                     /note= "EGF-like domain"

Query Match            78.2%;  Score 489;  DB 22;   Length 115;
Best Local Similarity  75.7%;  Pred. No. 1e-38;
Matches   87;  Conservative  12;  Mismatches  16;  Indels   0;  Gaps   0;

Qy    1 MPTDHEEPCGPSHKSFCLNGGLCVYIPTIPSPFCRCVENYTGARCEEVFLPGSSIQTKSN 60
         |||||| ||||||:|:||||||||||||||||||: ||||||||||||| ||| :||
Db    1 mptdheqpcgpshrsfclnggicvyiptipspfccienytgarceevflpsssipsesn 60

Qy   61 LFEAFVALAVLVTLIIGAFYFLCRKGHFQRASSVQYDINLVETSSTSAHHSHEQH 115
         | ||:|||||:||||:::| ||:||||||:  |:|||:|:::||
Db   61 lsaafvvlavltlliaalclcrkghlqrassvqceislsvetnnrtrhshreh 115
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

nucleic - nucleic search, using sw model

on: January 8, 2002, 15:14:05 ; Search time 1410.3 Seconds
(without alignments)
4070.778 Million cell updates/sec

le: US-09-640-041-3
fect score: 348
quence: 1 atgcaacagatcacgaaga.........acagtcatgaacacactga 348
ring table: IDENTITY_NUC Gapop 10.0 , Gapext 1.0
arched: 1472140 seqs, 824589755 residues
tal number of hits satisfying chosen parameters: 2944280
aimum DB seq length: 0
ximum DB seq length: 200000000
st-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

tabase :   GenEmbl:*

#	DB	#	DB
1:	gb_ba:*	19:	em_om:*
2:	gb_htg:*	20:	em_or:*
3:	gb_in:*	21:	em_ov:*
4:	gb_om:*	22:	em_pat:*
5:	gb_ov:*	23:	em_ph:*
6:	gb_pat:*	24:	em_pl:*
7:	gb_ph:*	25:	em_ro:*
8:	gb_pl:*	26:	em_sts:*
9:	gb_pr:*	27:	em_sy:*
10:	gb_ro:*	28:	em_un:*
11:	gb_sts:*	29:	em_vi:*
12:	gb_sy:*	30:	em_htgo_hum:*
13:	gb_un:*	31:	em_htgo_inv:*
14:	gb_vi:*	32:	em_htgo_rod:*
15:	em_ba:*	33:	em_htg_hum:*
16:	em_fun:*	34:	em_htg_inv:*
17:	em_hum:*	35:	em_htg_rod:*
18:	em_in:*	36:	em_htg_other:*

Result No.	Score	Query Match	Length	DB	ID	Description
1	348	100.0	348	6	AX088172	AX088172 Sequence
2	252	72.4	750	6	AX088170	AX088170 Sequence
3	252	72.4	750	10	AF083067	Mus muscu
4	157.8	45.3	166352	2	AC026180	Homo sapi
5	157.8	45.3	171000	2	AC087456	Homo sapi
6	157.8	45.3	189076	2	AC068838	Homo sapi
7	101.2	29.1	154323	2	AC026033	Homo sapi
8	101.2	29.1	161262	2	AC087464	Homo sapi
9	101.2	29.1	166352	2	AC026180	Homo sapi
10	90	25.9	180795	2	AC027104	Homo sapi
11	48.6	14.0	188553	2	AC090168	Homo sapi
12	37.2	10.7	204779	4	CNS01RH3	Homo sapi
13	35.4	10.2	2691	2	AC073727	Mus muscu
14	35.4	10.2	170139	2	AL136220	Human DNA
15	35	10.1	155084	2	AC073721	Mus muscu
16	35	10.1	190577	9	AC050036	Homo sapi
17	35	10.1	199684	9	AC079564	Mus muscu
18	34.6	10.1	241714	5	AF108420	Fugu rubr
19	34.6	10.0	46626	2	AC073373	Mus muscu
20	34.6	9.9	154242	2	AL158206	Human DNA
21	34.2	9.9	1121	2	AB023558	Human ade
22	34.2	9.8	1282	9	AC073248	Homo sapi
23	34.2	9.8	128294	9	AC018733	Homo sapi
24	34.2	9.8	163117	8	SPU73044	Schizosacch
25	33.8	9.7	2855	8	SPBC1861	S.pombe c
26	33.8	9.7	23851	8	AC044819	Homo sapi
27	33.8	9.7	151798	2	AC069199	Homo sapi
28	33.6	9.7	159991	2	AL162734	Homo sapi
29	33.6	9.7	236810	8	AL033391	C.albican
30	33.4	9.6	37968	2	CAC20C1	Homo sapi
31	33.4	9.6	75382	2	AC023242	Homo sapi
32	33.2	9.5	224863	2	AC093165	Homo sapi
33	33	9.5	177767	2	AC007342	Homo sapi
34	33	9.5	179202	2	AC025935	Homo sapi
35	33	9.5	188863	9	AC007345	Homo sapi
36	33	9.5	190063	2	AC023597	Homo sapi
37	32.8	9.4	17867	2	AC092170	Drosophil
38	32.8	9.4	36654	2	AC017551	Drosophil
39	32.8	9.4	126908	2	AC012365	Homo sapi
40	32.8	9.4	139515	2	AP002432	Homo sapi
41	32.8	9.4	159804	2	AC025876	Homo sapi
42	32.8	9.4	169152	9	AP000793	Homo sapi
43	32.8	9.4	177740	3	AC005714	Drosophil
44	32.8	9.4	181893	2	AC021901	Homo sapi
45	32.8	9.4	191466	3	AC008348	Drosophil

RESULT 1
AX088172

```
LOCUS       AX088172        348 bp    DNA        PAT    17-MAR-2001
DEFINITION  Sequence 3' from Patent WO0114415.
ACCESSION   AX088172
VERSION     AX088172.1    GI:13397084
KEYWORDS    .
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 348)
AUTHORS     Kavanaugh,W.M., Cen,H. and Lee,P.
TITLE       Egfh2 genes and gene products
JOURNAL     Patent: WO 0114415-A 3 01-MAR-2001;
            CHIRON CORPORATION (US)
FEATURES             Location/Qualifiers
source          1..348
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
```

ALIGNMENTS

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

RESULT 2
US-09-126-121-2
Sequence 2, Application US/09126121
Patent No. 6252051
GENERAL INFORMATION:
  APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
  TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
                      Ligands and Uses Therefor
  NUMBER OF SEQUENCES: 23
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Genentech, Inc.
    STREET: 1 DNA Way
    CITY: South San Francisco
    STATE: California
    COUNTRY: USA
    ZIP: 94080
  COMPUTER READABLE FORM:
    MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: WinPatin (Genentech)
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/09/126,121
    FILING DATE: 30-Jul-1998
    CLASSIFICATION:
  ATTORNEY/AGENT INFORMATION:
    NAME: Conley, Deirdre L.
    REGISTRATION NUMBER: 36,487
    REFERENCE/DOCKET NUMBER: P1084R1D1
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 650/225-2066
    TELEFAX: 650/952-9881
  INFORMATION FOR SEQ ID NO: 2:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 713 amino acids
      TYPE: Amino Acid
      TOPOLOGY: Linear
    FEATURE:
      NAME/KEY: Mouse NRG3 (mNRG3)/amino acid seq.
      LOCATION: 1-713
      IDENTIFICATION METHOD:
      OTHER INFORMATION:
US-09-126-121-2

Query Match   22.7%; Score 142; DB 4; Length 713;
Best Local Similarity 26.6%; Pred. No. 6.1e-08;
Matches 38; Conservative 20; Mismatches 47; Indels 38; Gaps 6;

```
Qy    3  TDHEEPCGPSHKSFCLNGGLCYVIPTIPS--PFCRCVENYTGARCEEVLP------ 51
         :|| .||  |  |||||.||:|| |::: | :||||   ||| | |||  : |||
Db  286  SEHFKPCRDKDLAYCLNDGECFVIEILTGSHKHCRCKEGYQGVRCDQ-FLPKTDSILSDP 344

Qy   52  ---GSSIQTKSNLFE------AFVALAVLVTLLIIGAFYFLCRKGHFQRASSVQ----- 95
         :|| :|| |   |        |:| || || |||| |||| || |: |||||
Db  345  TDHLGIEFMESEDVYQRQVLSISCIIFGIVIGVGMFCAAFYFKSKK----QAKQIQEHLKE 400

Qy   96  -----YDINLVETSTSAHHSH 112
               ||  ||  ||
Db  401  SQNGKNYSLKASSTKSESLMKSH 423
```

RESULT 3
US-08-168-091A-2
Sequence 2, Application US/08168091A
Patent No. 5665862
GENERAL INFORMATION:
  APPLICANT: Fischbach, Gerald.
  APPLICANT: Falls, Douglas R.
  APPLICANT: Rosen, Kenneth M.
  APPLICANT: Corfas, Gabriel.
  TITLE OF INVENTION: Neurotrophic Factor
  NUMBER OF SEQUENCES: 47
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: LAHIVE AND COCKFIELD
    STREET: 60 State Street, Suite 510
    CITY: Boston
    STATE: MA
    COUNTRY: USA
    ZIP: 02109
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: ASCII text
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/168,091A
    FILING DATE: 15-DEC-1993
    CLASSIFICATION: 435.
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 07/953,742
    FILING DATE: 29-SEP-1992
  ATTORNEY/AGENT INFORMATION:
    NAME: DeConti, Giulio A
    REGISTRATION NUMBER: 31,503
    REFERENCE/DOCKET NUMBER: HMI-002CP
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (617) 227-7400
    TELEFAX: (617) 227-5941
  INFORMATION FOR SEQ ID NO: 2:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 602 amino acids
      TYPE: amino acid
      TOPOLOGY: linear
    MOLECULE TYPE: protein
US-08-168-091A-2

Query Match   22.4%; Score 140; DB 1; Length 602;
Best Local Similarity 27.1%; Pred. No. 8.3e-08;
Matches 35; Conservative 21; Mismatches 51; Indels 22; Gaps 4;

```
Qy    3  TDHEEPCGPSHKSFCLNGGLCYVIPTIPSP---FCRCVENYTGARCEEV---------- 48
         :|| .||     |  ||:| |:  :| : |    |||||||  |||||
Db  135  TSHLTKCDIKQKAFCVNGGECMVKDLPNPPRYLCRCPNEFTGDRCQNYVMASFYKHLGI 194

Qy   49  -FLPGGSSIQTKSNLFEAFVALAVLVTLLIIGAFYFLC------RKGHFQRASSVQYDINLV 101
          ||  |:|| |  |     |:| |||| |:|| | |||        |||||||
Db  195  EFMEAEELYQKRVLTITGICIRALLVGIMCVVAY-CKTKKQRKKLHDRLRQSLRSERNNV 253

Qy  102  ETSSTSAHH 110
            |  |
Db  254  MNMANGPHH 262
```

RESULT 4
US-08-428-926-5
Sequence 5, Application US/08428926
Patent No. 5667780
GENERAL INFORMATION:
  APPLICANT: Ho, Wei-Hsien
  APPLICANT: Osheroff, Phyllis L.
  TITLE OF INVENTION: SENSORY AND MOTOR NEURON DERIVED FACTOR (SMDF)
  NUMBER OF SEQUENCES: 5